

TEKNIK PEMBELAJARAN MENDALAM BAGI  
RAMALAN UNIVARIAT DAN MULTIVARIAT  
UNTUK PASARAN SAHAM MALAYSIA

MOHD.RIDZUAN BIN AB.KHALIL

UNIVERSITI KEBANGSAAN MALAYSIA

TEKNIK PEMBELAJARAN MENDALAM BAGI RAMALAN UNIVARIAT DAN  
MULTIVARIAT UNTUK PASARAN SAHAM MALAYSIA

MOHD.RIDZUAN BIN AB.KHALIL

TESIS YANG DIKEMUKAKAN UNTUK MEMPEROLEH  
IJAZAH SARJANA SAINS DATA

FAKULTI TEKNOLOGI DAN SAINS MAKLUMAT  
UNIVERSITI KEBANGSAAN MALAYSIA  
BANGI

2022

**PENGAKUAN**

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

03 Jun 2022

MOHD.RIDZUAN BIN AB.KHALIL  
P106797

## PENGHARGAAN

Dengan nama Allah yang Maha Pemurah lagi Maha Mengasihani. Setinggi-tinggi kesyukuran dipanjatkan ke hadrat Allah S.W.T kerana dengan izin dan limpah kurnianya maka dapat saya menyiapkan dan menyempurnakan tesis ini dengan jayanya.

Jutaan terima kasih dan setinggi penghargaan diucapkan kepada penyelia saya Profesor Dr. Azuraliza Abu Bakar yang banyak memberikan bimbingan, tunjuk ajar dan dorongan sepanjang proses menyiapkan tesis ini. Segala pandangan, nasihat dan ilmu pengetahuan yang telah dicurahkan amatlah bernilai dan sangat membantu saya untuk menyiapkan tesis ini dengan sebaiknya. Ucapan terima kasih tidak terhingga juga kepada semua pensyarah Fakulti Teknologi dan Sains Maklumat (FTSM) UKM khususnya dari program Sarjana Sains Data yang telah mencurahkan ilmu pengetahuan dan bimbingan sepanjang pengajian saya.

Sekalung penghargaan kepada Jabatan Perkhidmatan Awam (JPA) yang telah memberikan saya peluang untuk melanjutkan pengajian secara sepenuh masa dan menaja serta membiayai sepenuhnya pengajian saya di UKM ini. Tidak dilupakan juga kepada semua rakan sepengajian yang sentiasa memberikan sokongan moral, idea serta berkongsi pendapat sepanjang pengajian ini khususnya untuk menyiapkan tesis ini.

Terima kasih khas kepada ahli keluarga saya yang sentiasa menyokong dan memberikan semangat kepada saya untuk meneruskan pengajian ini. Terima kasih juga buat semua yang terlibat sama ada secara langsung atau tidak langsung dalam memberikan sokongan, bantuan serta sumbangan sepanjang saya menyiapkan projek ini. Semoga Allah memberikan balasan yang berlipat ganda atas semua jasa kalian.

## ABSTRAK

Pasaran saham merupakan entiti institusi kewangan yang menjadi indikator dan memberi gambaran pertumbuhan serta kestabilan ekonomi sesebuah negara. Ia juga mempengaruhi dan memberi kesan secara tidak langsung kepada kelestarian perkembangan ekonomi yang bergantung kepada pasaran saham yang kukuh. Pasaran saham yang stabil mampu menarik pelabur untuk mengurus niaga dan melabur bagi mendapatkan pulangan keuntungan yang sewajarnya. Namun begitu, pelaburan di dalam pasaran saham adalah berisiko disebabkan kesukaran untuk memahami dan meramal pergerakan harganya. Disebabkan itu, keperluan kepada satu model ramalan yang lasak dan mampu memberikan keputusan yang tepat adalah penting kepada pelabur. Pada masa kini, pembelajaran mendalam telah menjadi pilihan utama untuk digunakan dalam pelbagai kajian bagi meramal pasaran saham disebabkan keupayaannya yang boleh memahami data tidak linear dan rumit. Kajian mengenai ramalan pasaran saham di Malaysia khususnya masih bersifat terpencil dan tidak begitu berkembang. Ini termasuk kajian penggunaan pembelajaran mendalam bagi ramalan pasaran saham. Oleh itu, kajian ini dilaksanakan untuk mencadangkan dan membangunkan satu model pembelajaran mendalam bagi meramal harga tutup saham di pasaran saham Malaysia. Sebanyak dua jenis data saham dari Bursa Malaysia yang tersenarai di dalam Indeks Komposit Kuala Lumpur (KLCI) iaitu Axiata Group Berhad dan Petronas Gas Berhad digunakan sebagai data kajian untuk membangunkan model ramalan tersebut. Sebanyak tiga algoritma pembelajaran mendalam iaitu Multilayer Perceptron (MLP), Rangkaian Neural Berlingkaran (CNN) dan Ingatan Jangka Pendek Panjang (LSTM) digunakan untuk membangunkan model ramalan. Tiga model regresi menggunakan pembelajaran mesin iaitu Regresi Pokok Keputusan (DTR), Regresi Hutan Rawak (RFR), dan Regresi Sokongan Vektor (SVR) turut dibangunkan sebagai model dasar untuk tujuan perbandingan prestasi ramalan. Hasil uji kaji menunjukkan ramalan univariat mempunyai sedikit kelebihan dari segi ketepatan, namun ramalan multivariat masih kompetitif dan menyimpan lebih banyak maklumat daripada set data. Perbandingan juga menunjukkan ketiga-tiga model pembelajaran mendalam mendapat prestasi paling baik dan mengatasi ketiga-tiga model pembelajaran mesin bagi ramalan untuk tempoh jangka pendek dan panjang. Model LSTM merupakan model terbaik bagi ramalan pasaran saham Malaysia bagi tempoh jangka panjang dan pendek dengan memperolehi ralat paling minimum berbanding lain.

## **DEEP LEARNING TECHNIQUE FOR MALAYSIAN STOCK MARKET PREDICTION**

### **ABSTRACT**

As part of a financial institution, stock market has been an important factor in the growth and stability of the national economy. It has indirect influences on the sustainability of economic growth which depends on a strong stock market. A stable stock market will ensure attraction to the investor to be part of the market and to get the return from their investment. Investment in stock market is risky because of the complexity and unpredictable nature of its price. Hence, there is a need to have a robust and accurate prediction model for the investor as a decision support tool towards the investment in stock market. Nowadays, deep learning has been the most chosen approach used in stock market prediction because of its ability to learn the non-linearity and complexity of stock market data. There is not much research and study done related to stock market prediction in Malaysia especially using the deep learning method. Therefore, this study aims to propose and develops a deep learning model that will predict the close price of Malaysian stock market. Axiata Group Berhad and Petronas Gas Berhad from Bursa Malaysia and listed in Kuala Lumpur Composite Index (KLCI) are used as datasets to develop the prediction model. Three different deep learning algorithms which are Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), and Long Short Term Memory (LSTM) are used to develop the prediction model. As a baseline model and performance comparison purposes, three additional regression models are developed using the traditional machine learning method which are Decision Tree Regression (DTR), Random Forest Regression (RFR), and Support Vector Regression (SVR). Experimental result shows that all three deep learning models achieve the highest accuracy and outperform all the three machine learning model in short and long-term prediction. It also shows that LSTM is the best prediction model for Malaysian stock market achieving the lowest prediction error among the other models.

## KANDUNGAN

		<b>Halaman</b>
<b>PENGAKUAN</b>		<b>ii</b>
<b>PENGHARGAAN</b>		<b>iii</b>
<b>ABSTRAK</b>		<b>iv</b>
<b>ABSTRACT</b>		<b>v</b>
<b>KANDUNGAN</b>		<b>vi</b>
<b>SENARAI JADUAL</b>		<b>ix</b>
<b>SENARAI ILUSTRASI</b>		<b>xii</b>
<b>SENARAI SINGKATAN</b>		<b>xv</b>
<b>BAB I</b>	<b>Pengenalan</b>	
1.1	Latar Belakang	1
	1.1.1 Bursa Saham Di Malaysia	3
	1.1.2 Ramalan Pasaran Saham	4
	1.1.3 Data Siri Masa	6
	1.1.4 Pendekatan Pembelajaran Mesin dan Pembelajaran Mendalam	7
1.2	Permasalahan Kajian	8
1.3	Objektif Kajian	10
	1.3.1 Objektif Umum	10
	1.3.2 Objektif Khusus	10
1.4	Skop Kajian	11
1.5	Metodologi Kajian	12
1.6	Kepentingan Kajian	13
1.7	Organisasi Tesis	13
<b>BAB II</b>	<b>KAJIAN KEPUSTAKAAN</b>	
2.1	Pengenalan	15
2.2	Latar Belakang Analisis Pasaran Saham	15
2.3	Ramalan Pasaran saham	18
2.4	Teknik Ramalan Pasaran saham	20
	2.4.1 Pendekatan Statistik Dan Model Linear	20
	2.4.2 Pembelajaran Mesin	22

2.5	Pembelajaran Mendalam Dalam Ramalan Pasaran Saham	25
2.5.1	Multilayer Perceptron (MLP)	27
2.5.2	Ingatan Jangka Pendek Panjang (LSTM)	28
2.5.3	Rangkaian Neural Berlingkaran (CNN)	29
2.6	Pemilihan Data Saham Dan Fitur Input Model Ramalan	31
2.6.1	Pemilihan Data	31
2.6.2	Pemilihan Fitur Input	33
2.7	Kajian Ramalan Pasaran saham	36
2.8	Kesimpulan	47
<b>BAB III</b>	<b>METODOLOGI KAJIAN</b>	
3.1	Pengenalan	49
3.2	Pemahaman Bisnes	50
3.3	Pemahaman Data	51
3.3.1	Data Axiata Group Berhad	53
3.3.2	Data Petronas Gas Berhad	55
3.4	Penyediaan Data	56
3.4.1	Pembersihan Data	56
3.4.2	Pengurangan Data	57
3.4.3	Transformasi Data	58
3.4.4	Penyediaan Set Data	61
3.5	Pemodelan	62
3.5.1	Pecahan Data Latihan dan Pengujian	63
3.5.2	Pembangunan Model Ramalan	64
3.6	Penilaian	70
3.6.1	Punca Min Ralat Kuasa Dua (RMSE)	70
3.6.2	Min Ralat Mutlak (MAE)	71
3.6.3	Min Peratus Ralat Mutlak (MAPE)	71
3.7	Kesimpulan	72
<b>BAB IV</b>	<b>KEPUTUSAN DAN PENILAIAN</b>	
4.1	Pengenalan	73
4.2	Ramalan Univariat	73
4.2.1	Ramalan Harga Saham Satu Hari Ke Depan	74
4.2.2	Ramalan Harga Saham Tiga Hari Ke Depan	79
4.2.3	Ramalan Harga Saham Lima Hari Ke Depan	85
4.2.4	Ramalan Harga Saham Tujuh Hari Ke Depan	91
4.3	Ramalan Multivariat	96
4.3.1	Ramalan Harga Saham Satu Hari Ke Depan	97



4.3.2	Ramalan Harga Saham Tiga Hari Ke Depan	102
4.3.3	Ramalan Harga Saham Lima Hari Ke Depan	107
4.3.4	Ramalan Harga Saham Tujuh Hari Ke Depan	112
4.4	Kesimpulan	117
<b>BAB V</b>	<b>ANALISIS DAN DAPATAN KAJIAN</b>	
5.1	Pengenalan	119
5.2	Penetapan Saiz Tetingkap	119
5.3	Ralat Ketepatan Ramalan Univariat dan Multivariat	121
5.3.1	Perbandingan Ralat Ketepatan Melalui Pengujian Statistik	123
5.4	Model Ramalan Terbaik	125
5.5	Kelebihan Algoritma LSTM Untuk Ramalan Pasaran Saham	130
5.6	Kesimpulan	132
<b>BAB VI</b>	<b>RUMUSAN</b>	
6.1	Pengenalan	133
6.2	Rumusan Kajian	133
6.3	Sumbangan Kajian	135
6.4	Cadangan Peluasan Kajian Di Masa Hadapan	136
6.5	Penutup	136
<b>RUJUKAN</b>		<b>137</b>
Lampiran A	Carta Perbandingan Ramalan Univariat Terbaik Bagi Model LSTM Berbanding Harga Sebenar	142
Lampiran B	Carta Perbandingan Ramalan Multivariat Terbaik Bagi Model LSTM Berbanding Harga Sebenar	145

## SENARAI JADUAL

<b>No. Jadual</b>		<b>Halaman</b>
Jadual 2.1	Ringkasan kajian ramalan pasaran saham	44
Jadual 3.1	Senarai fitur bagi set data kajian dan penerangan ringkas mengenainya	53
Jadual 3.2	Ringkasan penerangan statistik data Axiata Group Berhad	54
Jadual 3.3	Ringkasan penerangan statistik data Petronas Gas Berhad	55
Jadual 3.4	Set data serta variasi tetapan yang digunakan	62
Jadual 3.5	Peratusan pecahan set data serta jumlah rekod yang akan digunakan bagi tujuan latihan dan pengujian	63
Jadual 3.6	Bilangan model bagi setiap algoritma untuk kesemua tempoh ramalan	64
Jadual 3.7	Konfigurasi parameter bagi pembangunan model DTR	66
Jadual 3.8	Konfigurasi parameter bagi pembangunan model RFR	66
Jadual 3.9	Konfigurasi parameter bagi pembangunan model SVR	67
Jadual 3.10	Konfigurasi parameter bagi pembangunan model MLP	68
Jadual 3.11	Konfigurasi parameter bagi pembangunan model CNN	69
Jadual 3.12	Konfigurasi parameter bagi pembangunan model LSTM	69
Jadual 4.1	Analisis prestasi ramalan univariat dengan saiz tettingkap 10 hari	74
Jadual 4.2	Analisis prestasi ramalan univariat dengan saiz tettingkap 30 hari	75
Jadual 4.3	Analisis prestasi ramalan univariat dengan saiz tettingkap 60 hari	76
Jadual 4.4	Analisis prestasi ramalan harga saham 3 hari ke depan dengan saiz tettingkap 10 hari	80
Jadual 4.5	Analisis prestasi ramalan harga saham 3 hari ke depan dengan saiz tettingkap 30 hari	81
Jadual 4.6	Analisis prestasi ramalan harga saham 3 hari ke depan dengan saiz tettingkap 60 hari	82
Jadual 4.7	Analisis prestasi ramalan harga saham 5 hari ke depan dengan saiz tettingkap 10 hari	86

Jadual 4.8	Analisis prestasi ramalan harga saham 5 hari ke depan dengan saiz tettingkap 30 hari	86
Jadual 4.9	Analisis prestasi ramalan harga saham 5 hari ke depan dengan saiz tettingkap 60 hari	87
Jadual 4.10	Analisis prestasi ramalan harga saham 7 hari ke depan dengan saiz tettingkap 10 hari	91
Jadual 4.11	Analisis prestasi ramalan harga saham 7 hari ke depan dengan saiz tettingkap 30 hari	92
Jadual 4.12	Analisis prestasi ramalan harga saham 7 hari ke depan dengan saiz tettingkap 60 hari	93
Jadual 4.13	Analisis prestasi ramalan multivariat 1 hari ke depan dengan saiz tettingkap 10 hari	97
Jadual 4.14	Analisis prestasi ramalan multivariat 1 hari ke depan dengan saiz tettingkap 30 hari	98
Jadual 4.15	Analisis prestasi ramalan multivariat 1 hari ke depan dengan saiz tettingkap 60 hari	99
Jadual 4.16	Analisis prestasi ramalan multivariat 3 hari ke depan dengan saiz tettingkap 10 hari	102
Jadual 4.17	Analisis prestasi ramalan multivariat 3 hari ke depan dengan saiz tettingkap 30 hari	103
Jadual 4.18	Analisis prestasi ramalan multivariat 3 hari ke depan dengan saiz tettingkap 60 hari	104
Jadual 4.19	Analisis prestasi ramalan multivariat 5 hari ke depan dengan saiz tettingkap 10 hari	108
Jadual 4.20	Analisis prestasi ramalan multivariat 5 hari ke depan dengan saiz tettingkap 30 hari	108
Jadual 4.21	Analisis prestasi ramalan multivariat 5 hari ke depan dengan saiz tettingkap 60 hari	109
Jadual 4.22	Analisis prestasi ramalan multivariat 7 hari ke depan dengan saiz tettingkap 10 hari	113
Jadual 4.23	Analisis prestasi ramalan multivariat 7 hari ke depan dengan saiz tettingkap 30 hari	113
Jadual 4.24	Analisis prestasi ramalan multivariat 7 hari ke depan dengan saiz tettingkap 60 hari	114

Jadual 5.1	Penetapan saiz tettingkap yang mendapat prestasi ramalan terbaik bagi setiap tempoh ramalan untuk saham Axiata Group	120
Jadual 5.2	Penetapan saiz tettingkap yang mendapat prestasi ramalan terbaik bagi setiap tempoh ramalan untuk saham Petronas Gas	120
Jadual 5.3	Purata perbezaan nilai RMSE, MAE dan MAPE di antara ramalan univariat dan multivariat bagi kedua-dua saham	123
Jadual 5.4	Keputusan pengujian statistik bagi perbandingan ramalan univariat dengan multivariat untuk kesemua tempoh ramalan	124
Jadual 5.5	Kategori model yang mendapat prestasi ramalan terbaik bagi setiap tempoh ramalan untuk kedua-dua saham.	126
Jadual 5.6	Kedudukan tiga model terbaik bagi setiap tempoh ramalan untuk kedua-dua saham	127
Jadual 5.7	Perbandingan keputusan model LSTM kajian dengan beberapa kajian lepas yang berkaitan bagi ramalan harga saham satu hari ke depan	128
Jadual 5.8	Perbandingan keputusan model LSTM kajian dengan beberapa kajian lepas yang berkaitan bagi ramalan harga saham untuk tempoh jangka panjang	129
Jadual 5.9	Perbandingan prestasi ramalan yang diperolehi oleh model LSTM kajian ini dengan beberapa kajian pasaran saham Malaysia yang lepas	130
Jadual 5.10	Nilai RMSE, MAE dan MAPE yang diperolehi oleh model LSTM yang terbaik bagi setiap tempoh ramalan	132

## SENARAI ILUSTRASI

<b>No. Rajah</b>		<b>Halaman</b>
Rajah 2.1	Contoh struktur dan reka bentuk pembelajaran mendalam	26
Rajah 3.1	Metodologi CRISP-DM serta kitar hayat fasa di dalamnya	50
Rajah 3.2	Contoh data mentah Axiata Group Berhad	54
Rajah 3.3	Contoh data mentah Petronas Gas Berhad	56
Rajah 3.4	Samakan nilai yang hilang pada set data kajian	57
Rajah 3.5	Contoh set data selepas fitur yang tidak berkenaan dihapuskan	58
Rajah 3.6	Kaedah tettingkap bagi menukar data siri masa kepada set data dalam format pembelajaran berselia	61
Rajah 3.7	Proses pembangunan model-model ramalan	65
Rajah 4.1	Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Axiata Group	77
Rajah 4.2	Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Petronas Gas	78
Rajah 4.3	Ramalan model LSTM dengan pendekatan univariat ke atas harga tutup saham Axiata Group untuk tempoh satu hari ke depan berbanding harga sebenar	79
Rajah 4.4	Ramalan model LSTM dengan pendekatan univariat ke atas harga tutup saham Petronas Gas untuk tempoh satu hari ke depan berbanding harga sebenar	79
Rajah 4.5	Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Axiata Group	83
Rajah 4.6	Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Petronas Gas	84
Rajah 4.7	Ramalan model LSTM dengan pendekatan univariat ke atas harga tutup saham Axiata Group untuk tempoh tiga hari ke depan berbanding harga sebenar	85
Rajah 4.8	Ramalan model LSTM dengan pendekatan univariat ke atas harga tutup saham Petronas Gas untuk tempoh tiga hari ke depan berbanding harga sebenar	85

Rajah 4.9	Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Axiata Group	89
Rajah 4.10	Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Petronas Gas	89
Rajah 4.11	Ramalan model LSTM dengan pendekatan univariat ke atas harga tutup saham Axiata Group untuk tempoh lima hari ke depan berbanding harga sebenar	90
Rajah 4.12	Ramalan model MLP dengan pendekatan univariat ke atas harga tutup saham Petronas Gas untuk tempoh lima hari ke depan berbanding harga sebenar	91
Rajah 4.13	Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Axiata Group	94
Rajah 4.14	Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Petronas Gas	95
Rajah 4.15	Ramalan model MLP dengan pendekatan univariat ke atas harga tutup saham Axiata Group untuk tempoh tujuh hari ke depan berbanding harga sebenar	96
Rajah 4.16	Ramalan model MLP dengan pendekatan univariat ke atas harga tutup saham Petronas Gas untuk tempoh tujuh hari ke depan berbanding harga sebenar	96
Rajah 4.17	Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Axiata Group	100
Rajah 4.18	Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Petronas Gas	101
Rajah 4.19	Ramalan model LSTM dengan pendekatan multivariat ke atas harga tutup saham Axiata Group untuk tempoh satu hari ke depan berbanding harga sebenar	101
Rajah 4.20	Ramalan model LSTM dengan pendekatan multivariat ke atas harga tutup saham Petronas Gas untuk tempoh satu hari ke depan berbanding harga sebenar	102
Rajah 4.21	Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Axiata Group	105
Rajah 4.22	Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Petronas Gas	106
Rajah 4.23	Ramalan model LSTM dengan pendekatan multivariat ke atas harga tutup saham Axiata Group untuk tempoh satu hari ke depan berbanding harga sebenar	107

Rajah 4.24	Ramalan model LSTM dengan pendekatan multivariat ke atas harga tutup saham Petronas Gas untuk tempoh satu hari ke depan berbanding harga sebenar	107
Rajah 4.25	Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Axiata Group	111
Rajah 4.26	Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Petronas Gas	111
Rajah 4.27	Ramalan model LSTM dengan pendekatan multivariat ke atas harga tutup saham Axiata Group untuk tempoh satu hari ke depan berbanding harga sebenar	112
Rajah 4.28	Ramalan model MLP dengan pendekatan multivariat ke atas harga tutup saham Petronas Gas untuk tempoh satu hari ke depan berbanding harga sebenar	112
Rajah 4.29	Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Axiata Group	116
Rajah 4.30	Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Petronas Gas	116
Rajah 4.31	Ramalan model LSTM dengan pendekatan multivariat ke atas harga tutup saham Axiata Group untuk tempoh satu hari ke depan berbanding harga sebenar	117
Rajah 4.32	Ramalan model LSTM dengan pendekatan multivariat ke atas harga tutup saham Petronas Gas untuk tempoh satu hari ke depan berbanding harga sebenar	117
Rajah 5.1	Perbandingan nilai RMSE yang diperolehi oleh model terbaik bagi setiap tempoh ramalan yang menggunakan pendekatan univariat dan multivariat	122
Rajah 5.2	Perbandingan nilai MAE yang diperolehi oleh model terbaik bagi setiap tempoh ramalan yang menggunakan pendekatan univariat dan multivariat	122
Rajah 5.3	Perbandingan nilai MAPE yang diperolehi oleh model terbaik bagi setiap tempoh ramalan yang menggunakan pendekatan univariat dan multivariat	122

## SENARAI SINGKATAN

ANN	Rangkaian Neural Buatan
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
BSKL	Bursa Saham Kuala Lumpur
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
CNN	Rangkain Neural Berlingkaran
CPU	Unit Pemrosesan Pusat
CRISP-DM	Cross-Industry Standard Process for Data Mining
DAX	Deutscher Aktien Index
DRSA	Data Raya Sektor Awam
DT	Pokok Keputusan
DTR	Regresi Pokok Keputusan
ECG	Elektrokardiogram
EMH	Hipotesis Pasaran Cepak
EWS	Sistem Amaran Awal
EWT	Empirical Wavelet Transform
GA	Algoritma Genetik
GARCH	Generalized Autoregressive Conditional Heteroskedastic
GPU	Unit Pemrosesan Grafik
GRU	Unit Berulang Berpagar
HSI	Hang Seng Index
KLCI	Indeks Komposit Kuala Lumpur
KNN	K Jiran Terdekat



KOSPI	Indeks Komposit Korea
LR	Regresi Linear
LSTM	Ingatan Jangka Pendek Panjang
MACD	Moving Average Convergence/Divergence
MAE	Min Ralat Mutlak
MAPE	Min Peratus Ralat Mutlak
MLP	Multilayer Perceptron
NARX	Nonlinear Autoregressive Network With Exogenous Input
NSE	Nigeria Stock Exchange
NYSE	New York Stock Exchange
ORELM	Outlier Robust Extreme Learning Machine
PKS	Perusahaan Kecil Sederhana
RF	Hutan Rawak
RFR	Regresi Hutan Rawak
RMSE	Punca Min Ralat Kuasa Dua
RNN	Rangkaian Neural Berulang
SMA	Simple Moving Average
SSE	Shanghai Stock Exchange Composite Index
STAR	Smooth Transition Autoregressive
SVM	Mesin Sokongan Vektor
SVR	Regresi Sokongan Vektor

## **BAB I**

### **PENGENALAN**

#### **1.1 LATAR BELAKANG**

Pasaran saham merupakan salah satu institusi kewangan yang sangat penting dan mempunyai kesan langsung terhadap perkembangan ekonomi sesebuah negara. Ia juga merupakan salah satu indikator penentu bagi menggambarkan kestabilan ekonomi sesebuah negara. Pengaruh pasaran saham sama ada untuk tempoh jangka masa pendek atau panjang mampu merubah kadar pertumbuhan ekonomi dan pelaburan di dalam sesebuah negara dengan drastik dan menyeluruh (Masoud 2013; Azam et al. 2016). Atas faktor itu, keupayaan sesebuah negara untuk mengekalkan pasaran saham yang berdaya maju dan mampan adalah penting demi menjamin kestabilan ekonomi negara tersebut. Pasaran saham yang kukuh akan memiliki keupayaan untuk menarik minat pelabur dari dalam dan luar negara untuk membeli saham terbitan. Jumlah dagangan yang meningkat di dalam pasaran tersebut akan menambah dana dan peluang kepada syarikat berkaitan untuk memperluaskan ekonomi mereka dan secara tidak langsung mengukuhkan lagi pertumbuhan ekonomi negara terbabit. Kestabilan dan pertumbuhan ekonomi yang mampan turut memberi kesan ke atas perkembangan pasaran saham. Pradhan (2018) telah menghuraikan mengenai hubungan dan kesan dua hala di antara pasaran saham dan juga pertumbuhan ekonomi sesebuah negara di mana keduanya mempunyai kesan langsung antara satu sama lain. Disebabkan itu, kestabilan pasaran saham dan juga ekonomi negara sentiasa diberi perhatian khusus demi menjamin pertumbuhan dan perkembangan sesebuah negara.

Pasaran saham turut dikenali dengan nama pasaran stok atau pasaran ekuiti di mana ianya merupakan pasaran kewangan yang membolehkan pembeli dan penjual beroperasi dan bertransaksi dengan saham syarikat milikan awam. Ia juga merupakan

satu pasaran kewangan yang menjadi pusat kepada pertumbuhan dan perkembangan ekonomi yang kukuh dan berdaya saing (Masoud 2013). Pasaran saham membantu pertumbuhan ekonomi melalui pengumpulan dana jangka panjang daripada pelabur-pelabur untuk digunakan bagi program-program pembangunan fizikal negara. Ia juga menjadi platform kepada pelbagai syarikat untuk penyediaan modal dengan jumlah yang besar bagi memperluaskan kegiatan ekonomi mereka. Ini dapat mengelak syarikat daripada membuat hutang dan mengelakkan pertambahan perbelanjaan ke atas hutang tersebut. Saham merujuk kepada sekuriti yang merupakan sebahagian daripada modal pemilik dalam sesebuah perniagaan. Pemegang saham adalah individu atau institusi yang membeli saham tersebut dan menjadi pemilik bersama perniagaan berkenaan serta berkongsi keuntungan dan kerugian daripadanya.

Secara umumnya, mana-mana pihak boleh mengapungkan syarikat mereka secara awam dengan menjual sebahagian pegangan syarikat tersebut kepada pelabur individu atau institusi. Ini memberi peluang kepada syarikat tersebut untuk memperluaskan perniagaan dan menambah keupayaan kewangannya. Pembelian dan penjualan saham ini biasanya diuruskan oleh organisasi yang dikenali sebagai bursa saham di setiap negara. Di Malaysia, organisasi tersebut dikenali dengan nama Bursa Malaysia di mana ia bertanggungjawab untuk menyediakan platform pasaran dan perkhidmatan pelaburan serta pelbagai maklumat saham kepada pelabur. Melalui bursa saham, individu atau institusi dapat membeli dan mengaut keuntungan daripada pasaran saham melalui bayaran dividen atau penjualan semula saham dengan harga yang lebih mahal. Proses penjualan dan pembelian saham bagi tujuan mendapatkan keuntungan memerlukan maklumat dan analisis yang terperinci. Ini kerana ia melibatkan risiko kerugian yang besar sekiranya tidak dibuat berpandukan kepada maklumat dan analisis berkenaan. Maklumat-maklumat berkaitan saham berkenaan serta analisis berkaitannya sentiasa digunakan oleh pelabur dan ahli ekonomi dalam meramal pergerakan harga saham untuk tempoh jangka masa pendek atau panjang. Ini memberi panduan dan rujukan kepada mereka untuk mengambil langkah dan tindakan yang perlu dalam mendapatkan keuntungan di atas transaksi jual beli saham terbabit.

### 1.1.1 Bursa Saham Di Malaysia

Pasaran saham di Malaysia diuruskan oleh Bursa Malaysia selaku organisasi yang menyediakan kemudahan dan perkhidmatan kepada penjual dan pembeli untuk berurus niaga secara lebih efisien. Kewujudan organisasi ini dapat menjamin suatu keadaan dan persekitaran pasaran yang terbuka serta telus dalam berurus niaga serta mampu memberi keyakinan kepada pelabur untuk terlibat dalam urus niaga di sini (Hafizah & Saiful, 2013). Bursa Malaysia merupakan syarikat yang diletakkan di bawah kawal selia Suruhanjaya Sekuriti dan juga Kementerian Kewangan Malaysia. Terdapat beberapa akta yang diguna pakai bagi tujuan kawal selia pasaran modal ke atas Bursa Malaysia seperti Akta Pasaran Modal dan Perkhidmatan 2007, Akta Suruhanjaya Sekuriti 1993, Akta Syarikat 1965, Akta Perindustrian Sekuriti 1991 dan lain-lain lagi. Sebagai satu pasaran yang inklusif, Bursa Malaysia mempunyai lebih daripada 900 syarikat yang disenaraikan melalui Pasaran Utama bagi syarikat yang bermodal besar, Pasaran ACE bagi syarikat yang terdiri dari pelbagai saiz dan Pasaran LEAP bagi syarikat Perusahaan Kecil dan Sederhana (PKS).

Bursa Malaysia sebelum ini dikenali dengan nama Bursa Saham Kuala Lumpur (BSKL) telah bermula sejak awal tahun 1930 bagi urusan saham di Malaysia. Namun dagangan saham awam hanya bermula pada awal tahun 1960 di mana Bursa Saham Malaya ditubuhkan. Kini Bursa Malaysia telah menjadi satu daripada bursa yang terbesar di rantau Asia Tenggara dan menjadi organisasi yang cemerlang dalam mengawal selia bursa saham bersepadu dan pelbagai perkhidmatan lain yang berkaitan. Pengiktirafan ini disokong melalui pelbagai anugerah dan persijilan di peringkat antarabangsa yang diterima oleh Bursa Malaysia setiap tahun.

Sebagai penanda aras dan indikator kekuatan ekonomi sesebuah negara, indeks pasaran saham dijadikan sebagai rujukan utama oleh pelabur (Hafizah & Saiful, 2013). Atas tujuan tersebut, Indeks Komposit Kuala Lumpur (KLCI) diperkenalkan pada tahun 1986 bagi menggambarkan kedudukan pasaran saham di Malaysia. Pada tahun 2009, Bursa Malaysia telah mengambil inisiatif untuk kekal relevan di peringkat global dengan menambah baik indeks pasaran saham ini menggunakan metodologi pengiraan indeks FTSE yang diterima pakai di peringkat antarabangsa. Penjenamaan semula nama KLCI turut dilakukan di mana ia dikenali dengan nama baharu iaitu FTSE Bursa

Malaysia KLCI. Ia menyenaraikan 30 buah syarikat di papan utama Malaysia di mana syarikat-syarikat ini adalah yang paling besar dan memenuhi keperluan dan syarat asas yang ditetapkan oleh FTSE Bursa Malaysia.

### **1.1.2 Ramalan Pasaran Saham**

Keupayaan memahami dan menganalisis pasaran saham adalah asas utama yang perlu ada pada setiap pelabur yang terlibat dengan urus niaga tersebut. Ini kerana urus niaga saham adalah transaksi yang berisiko tinggi disebabkan perubahan harga yang tidak dapat dijangka dan boleh membawa kepada kerugian yang besar. Ia juga boleh memberikan keuntungan berlipat kali ganda sekiranya jangkaan dan analisis yang tepat dibuat apabila harga saham yang dibeli meningkat daripada harga asalnya. Bagi mendapatkan rujukan dan panduan yang tepat untuk mengurus niaga, pelabur perlu memahami dan membuat analisis ke atas pasaran saham. Dua kaedah konvensional yang biasa digunakan dalam membuat jangkaan harga saham adalah melibatkan analisis fundamental dan analisis teknikal (Bustos & Pomares-Quimbaya 2020; Chung & Shin, 2020; Abu-Mostafa & Atiya, 1996). Dengan menggunakan dua kaedah ini, pelabur boleh membuat jangkaan corak dan juga pergerakan harga saham pada masa akan datang. Ia seterusnya dapat membantu pelabur membuat keputusan untuk melabur di dalam pasaran saham serta mengurangkan risiko kerugian. Namun begitu, kedua-dua analisis ini memerlukan masa, pemerhatian dan bacaan yang sangat rumit serta sukar untuk difahami dengan mudah.

Menurut Lee et al., (2021) analisis fundamental merujuk kepada penilaian ke atas faktor-faktor yang boleh mempengaruhi harga saham seperti faktor ekonomi, kewangan, keadaan industri, pasaran saham, politik dan pelbagai faktor yang lain bagi meramal dan menjangka harga sebenar saham tersebut. Ia melibatkan data yang banyak dan besar serta rumit untuk dianalisa. Lee et al. juga mendefiniskan analisis teknikal sebagai kaedah analisa yang menggunakan data harga saham yang lepas serta jumlah jualannya bagi meramal dan menjangka pergerakan harga saham. Kebiasaannya, kaedah ini akan menggunakan visual seperti carta dan graf bagi memudahkan analisis dibuat. Selain itu, beberapa indikator teknikal juga akan dijana menggunakan data harga saham tersebut. Di antara kedua-dua kaedah ini, analisis teknikal adalah kaedah yang

selalu digunakan oleh pelabur bagi membuat analisis dan meramal pergerakan pasaran saham sebelum keputusan untuk membeli atau menjual saham yang berkaitan dibuat.

Secara asasnya, pasaran saham dan perubahan harganya adalah sukar untuk dijangka dan diramal disebabkan oleh sifatnya yang sentiasa berubah dan mempunyai pergerakan turun dan naik yang bersifat rawak. Ia juga mempunyai sifat hingar, dinamik, tidak pegun dan tidak linear yang menyebabkan pergerakan harga saham ini sukar untuk diramal dan dijangka (Abu-Mostafa & Atiya 1996). Disebabkan itu, kadar ketepatan yang tinggi dalam membuat ramalan pasaran saham adalah satu cabaran yang sangat besar di dalam industri pelaburan. Walau bagaimanapun, isu-isu tersebut tidak menghalang dan menjadi rintangan untuk pelbagai pihak menjadikan ramalan pasaran saham sebagai satu kajian dan pelaksanaan sebenar sama ada di bidang akademik atau pun secara praktikal.

Penggunaan model linear yang mengadaptasikan pendekatan statistik adalah teknik konvensional yang digunakan untuk membuat ramalan ke atas pasaran saham sebelum wujudnya teknik yang lebih moden seperti pembelajaran mesin (Shah et al. 2019). Walaupun telah lama digunakan, ia masih mampu menandingi teknik-teknik moden seperti pembelajaran mesin dengan memberikan keputusan dan ramalan yang agak baik dan tepat khususnya bagi ramalan jangka pendek. Namun begitu, teknik ini mempunyai beberapa kekangan ke atas data siri masa seperti keperluan data yang pegun, linear dan tertabur normal. Ini menyebabkan ramai penyelidik mula beralih kepada teknik yang lebih moden seperti pembelajaran mesin khususnya teknik pembelajaran mendalam. Teknik ini terbukti mampu memberikan keputusan ramalan yang lebih baik dan tepat di mana model yang dibangunkan lebih tahan lasak dan fleksibel dalam mempelajari data siri masa pasaran saham.

Penglibatan di dalam pelaburan saham telah menjadi satu daripada cabang perniagaan yang semakin popular di Malaysia. Perkembangan teknologi maklumat dan infrastruktur berkaitannya telah menyumbang dan menjadi faktor terbesar kepada lonjakan perkara ini. Dengan meluasnya liputan internet dan pemilikan telefon pintar serta komputer peribadi di kalangan rakyat Malaysia, informasi serta capaian kepada platform pasaran saham menjadi lebih mudah dan cepat. Pelaburan saham kini bukan

lagi menjadi profesion khusus kepada pelabur tetapi turut dijadikan sebagai kerjaya dan sumber pendapatan sampingan pelbagai lapisan masyarakat di Malaysia.

Walaupun pelaburan di dalam pasaran saham menawarkan keuntungan dalam kadar yang tinggi, ia juga sebenarnya adalah satu urusan niaga yang sangat berisiko disebabkan ketidakpastian perubahan harga saham yang sukar untuk dijangka dan diramal. Pasaran saham juga bersifat dinamik dan dipengaruhi oleh pelbagai faktor sekeliling seperti ekonomi, politik, sentimen dan lain-lain lagi. Namun begitu, adalah sesuatu yang mustahil untuk seseorang pelabur menilai dan menganalisa kesemua faktor tersebut bagi membuat jangkaan dan ramalan harga saham. Pelabur juga perlu sentiasa peka dengan pergerakan pasaran saham dan memerlukan maklumat yang cepat dan tepat untuk membuat keputusan yang efektif dan mampu memberi pulangan kepadanya. Pada masa kini, analisis dan alat bantuan keputusan yang biasa digunakan oleh pelabur kebanyakannya hanya berasaskan kepada analisis fundamental dan analisis teknikal sahaja (Lee et al. 2021).

Berdasarkan kepada isu dan permasalahan tersebut, keperluan kepada satu model yang fleksibel, mampu memahami corak serta pergerakan pasaran saham dan seterusnya berupaya untuk meramal harga saham adalah sangat penting kepada pelabur. Ini akan membantu pelabur dalam membuat pertimbangan sewajarnya sebelum keputusan sebenar diambil serta membolehkan pelabur melabur dengan bijak dan tepat bagi mendapatkan faedah dan keuntungan yang optimum. Ia juga dapat membantu mengurangkan risiko kerugian disebabkan oleh penurunan harga saham yang tidak dapat dijangka.

### **1.1.3 Data Siri Masa**

Data siri masa adalah koleksi turutan data yang direkodkan pada setiap sela dan tempoh masa tertentu di mana setiap satu data tersebut mempunyai rekod masa ianya diceraip (Mahalakshmi et al. 2016). Data siri masa disusun mengikut kronologi masa ianya direkod dan kebiasaannya diterjemahkan ke dalam bentuk visual seperti graf dan carta. Dengan perkembangan teknologi seperti sensor dan sistem perkomputeran yang canggih, pelbagai jenis data siri masa boleh diperolehi di dalam pelbagai bidang seperti

perubatan, kewangan, meteorologi, ekonomi dan lain-lain lagi. Antara contoh data siri masa adalah seperti data elektrokardiogram (ECG), data cuaca, taburan hujan, perolehan, kadar pertukaran matawang dan juga termasuk data harga pasaran saham.

Perlombongan data merujuk kepada teknik dan proses menggali dan mendapatkan maklumat serta pengetahuan tersembunyi dari koleksi data yang besar dan banyak. Penggunaan pendekatan perlombongan data ke atas data siri masa kebiasaannya digunakan bagi mengenal pasti corak dan trend pergerakan data tersebut. Melaluinya, analisis prediktif boleh dibuat seperti klasifikasi pergerakan data dan juga ramalan nilai pada masa hadapan berdasarkan sejarah data tersebut. Bagi data pasaran saham, teknik ini digunakan untuk meramal turun naik harga saham dan nilai harga saham tersebut pada masa hadapan. Di dalam data siri masa, data pada masa yang lepas adalah keperluan utama dan paling penting bagi membuat ramalan masa hadapan (Al-Mashhadani et al. 2021). Pada masa kini, pendekatan pembelajaran mesin khususnya pembelajaran mendalam menjadi kaedah dan teknik yang popular digunakan untuk membuat perlombongan data ke atas data siri masa pasaran saham. Ia memberikan prestasi ketepatan ramalan yang tinggi khususnya bagi meramal harga dan juga pergerakan turun naiknya pada masa hadapan.

#### **1.1.4 Pendekatan Pembelajaran Mesin dan Pembelajaran Mendalam**

Perkembangan teknologi di era globalisasi dan digital kini bergerak dengan begitu pantas di mana pelbagai inovasi baharu diperkenalkan khususnya di dalam bidang kecerdasan buatan. Pembelajaran mesin merupakan salah satu cabang utama di dalam bidang kecerdasan buatan dan berkembang dengan begitu pesat. Teknologi ini mampu meniru keupayaan manusia untuk belajar secara sendiri dan menjadi pintar menggunakan data dan algoritma yang telah ditetapkan tanpa perlu diprogramkan. Penggunaan pembelajaran mesin mampu memudahkan pelbagai tugas dan meningkatkan produktiviti dalam pelbagai bidang. Melalui data yang diberikan kepada model pembelajaran mesin, ia mampu belajar daripadanya dan seterusnya membuat analisis prediktif. Negara Malaysia turut tidak ketinggalan untuk berada di atas landasan teknologi ini demi membolehkan kita terus berdaya saing dan kompeten di dalam era digital. Ia juga bagi memastikan teknologi ini mampu meningkatkan kecekapan tahap penyampaian perkhidmatan Kerajaan kepada rakyat dan seterusnya memberi manfaat



secara lebih inklusif. Selain itu, ia mampu mengurangkan kos operasi organisasi dalam membuat perancangan serta keputusan yang lebih tepat. Selaras dengan hasrat Malaysia untuk meningkatkan kedudukan negara sebagai Kerajaan Digital, Projek Perkhidmatan Analitis Data Raya Sektor Awam (DRSA) 2.0 telah dilancarkan pada 10 November 2021 di mana pembelajaran mesin merupakan salah satu inisiatif utama di bawah program tersebut.

Pembelajaran mendalam pula adalah satu evolusi daripada pembelajaran mesin dan dibangunkan berasaskan kepada algoritma Rangkaian Neural Buatan (ANN). Ia merupakan subset di dalam pembelajaran mesin dan menggunakan kaedah yang lebih kompleks serta moden untuk belajar secara sendiri dan membuat analisis prediktif yang lebih tepat tanpa perantaraan dan bantuan manusia. Ia cuba untuk meniru konsep otak manusia berfikir melalui kombinasi input data, pemberat dan kecenderungan yang ditetapkan di dalam algoritma. Kesemua elemen ini akan cuba mengenal pasti, mengelaskan dan memperincikan objek yang ada di dalam data secara tepat dan pintar. Pembelajaran mendalam mula digunakan secara meluas dalam pelbagai bidang seperti pengecaman imej (Pak & Kim 2018), kesihatan (Zeroual et al. 2020) dan juga kewangan (Livieris et al. 2020). Pada masa kini, banyak kajian tentang ramalan data siri masa juga turut menggunakan pembelajaran dalaman sebagai pendekatan baharu dalam membuat analisis prediktif (Ismail Fawaz et al. 2019). Ini termasuk ramalan ke atas pasaran saham khususnya bagi meramal pergerakan dan harga saham pada masa akan datang dengan lebih tepat. Namun begitu, kajian serta pembangunan model ramalan berkaitan pasaran saham di Malaysia masih kurang dan tidak begitu berkembang khususnya melalui pendekatan pembelajaran mendalam. Penggunaan pembelajaran mendalam juga mempunyai kemampuan ramalan yang terbukti mampu memberikan prestasi dan ketepatan yang baik berbanding model pembelajaran mesin tradisional (Rama Krishna et al. 2020; Sezer et al. 2020).

## **1.2 PERMASALAHAN KAJIAN**

Berdasarkan kepada huraian dan latar belakang yang dinyatakan di dalam tajuk 1.1 di atas, permasalahan kajian telah dikenal pasti berdasarkan kepada jurang dan isu yang timbul daripada ramalan pasaran saham di Malaysia. Isu utama adalah disebabkan kajian dan penyelidikan berkaitan ramalan pasaran saham di Malaysia yang dilihat

masih kurang khususnya berkaitan dengan penggunaan teknik pembelajaran mendalam. Ini jauh berbeza dengan kajian-kajian yang dibuat untuk pasaran saham luar Malaysia yang telah lama memberi tumpuan kepada pendekatan baharu ini. Perkara ini telah mewujudkan satu jurang yang perlu diisi melalui pelaksanaan kajian ini bagi meluaskan lagi kajian ramalan pasaran saham di Malaysia menggunakan teknik pembelajaran mendalam.

Selain itu, pelabur pasaran saham sentiasa memerlukan kepada peralatan dan perantara yang mampu membantu mereka meramal corak serta harga pasaran saham (Kim Soon et al. 2018). Disebabkan itu, keperluan kepada satu model ramalan yang lasak dan mampu memberikan keputusan dengan cepat dan tepat adalah penting kepada pelabur-pelabur di Malaysia. Ini disebabkan perubahan harga pasaran saham adalah sangat drastik dan di luar jangkaan pelabur (Al-Mashhadani et al. 2021). Model ramalan yang lasak mampu memberikan ramalan yang konsisten dan tepat walaupun berlaku perubahan pada fitur input atau harga lepas yang digunakan untuk meramal harga saham pada masa hadapan. Keputusan ramalan yang pantas serta kemampuan meramal bagi tempoh jangka panjang juga dapat membantu pelabur merancang serta menentukan tindakan ke atas pelaburan mereka. Penggunaan teknik pembelajaran mendalam sebagai asas kepada model ramalan mampu memberikan ciri-ciri yang dinyatakan untuk digunakan sebagai medium dan alat bantuan keputusan kepada para pelabur. Melalui kajian-kajian lepas juga, kebanyakan teknik pembelajaran mendalam yang digunakan didapati mampu memberikan prestasi ramalan yang lebih baik jika dibandingkan dengan teknik pembelajaran mesin tradisional. Oleh itu, kesinambungan kesemua kajian tersebut perlu diteruskan dengan melihat kepada skop dan pelaksanaannya di dalam pasaran saham Malaysia.

Hasil daripada permasalahan kajian yang telah dinyatakan, penyelesaian ke atasnya melalui kajian ini dilaksanakan untuk menjawab persoalan seperti berikut:

- a) Adakah teknik ramalan melalui pemilihan fitur input daripada data harga saham berdasarkan pendekatan univariat dan multivariat serta pemilihan bilangan fitur input melalui saiz tetingkap akan mempengaruhi prestasi ramalan pasaran saham Malaysia?

- b) Bagaimana ketepatan prestasi ramalan model pembelajaran mendalam jika dibandingkan dengan model pembelajaran mesin tradisional? Adakah ia mampu memberikan prestasi ramalan yang baik di dalam skop pasaran saham Malaysia?
- c) Apakah algoritma pembelajaran mendalam yang paling sesuai bagi pasaran saham Malaysia dan mampu memberikan prestasi yang baik ke atas model ramalan yang dibangunkan?

### **1.3 OBJEKTIF KAJIAN**

Kajian ini dilaksanakan untuk mencapai objektif yang dinyatakan seperti berikut:

#### **1.3.1 Objektif Umum**

Objektif umum kajian ini adalah untuk mencadangkan serta membangunkan model ramalan harga saham di Malaysia berasaskan kepada algoritma pembelajaran mendalam.

#### **1.3.2 Objektif Khusus**

Objektif khusus kajian ini dilaksanakan adalah seperti berikut:

- a) Membuat perbandingan prestasi ramalan di antara pendekatan univariat dan multivariat serta pemilihan saiz tettingkap yang ideal berdasarkan kepada pemilihan fitur input yang mempengaruhi prestasi dan ketepatan model ramalan;
- b) Membuat perbandingan prestasi ramalan di antara teknik pembelajaran mendalam dengan pembelajaran mesin tradisional bagi menentukan pendekatan yang mampu memberikan ketepatan yang lebih baik ke atas ramalan pasaran saham Malaysia; dan
- c) Menentukan model ramalan berasaskan algoritma pembelajaran mendalam yang terbaik bagi ramalan harga saham di Malaysia.

#### 1.4 SKOP KAJIAN

Skop kajian ini memberi fokus kepada cadangan pembangunan tiga model pembelajaran mendalam dan tiga model pembelajaran mesin bagi meramal harga saham untuk tempoh jangka pendek dan panjang. Bagi tujuan tersebut, data sekunder yang dimuat turun daripada laman sesawang Yahoo Finance akan digunakan bagi melatih dan menguji kesemua model tersebut. Dua jenis data saham di Malaysia dari FTSE Bursa Malaysia KLCI akan digunakan iaitu data saham Axiata Group Berhad dari sektor telekomunikasi dan data saham Petronas Gas Berhad dari sektor Gas dan Utiliti. Data-data ini diambil daripada tahun 2017 sehingga 2021 yang meliputi tempoh hampir 5 tahun.

Kajian ini merangkumi 2 uji kaji utama dalam meramal harga saham. Ia melibatkan ramalan jangka pendek iaitu ramalan harga tutup saham pada keesokan harinya dan ramalan jangka panjang iaitu ramalan pada hari ketiga, kelima dan ketujuh. Bagi model pembelajaran mendalam, algoritma digunakan adalah Rangkaian Neural Buatan (ANN), Rangkaian Neural Berlingkaran (CNN) dan Ingatan Jangka Pendek Panjang (LSTM) manakala Regresi Pokok Keputusan (DTR), Regresi Hutan Rawak (RFR) dan Regresi Sokongan Vektor (SVR) dipilih untuk model pembelajaran mesin. Hasil keputusan daripada semua model yang dibangunkan akan dinilai menggunakan tiga metrik penilaian iaitu Punca Min Ralat Kuasa Dua (RMSE), Min Ralat Mutlak (MAE) dan Min Peratus Ralat Mutlak (MAPE).

Bagi tujuan perbandingan prestasi, ketiga-tiga model pembelajaran mesin yang dibangunkan akan menjadi model dasar untuk melihat keupayaan dan kemampuan ketiga-tiga model pembelajaran mendalam untuk membuat ramalan. Analisis ke atas hasil keputusan akan dibuat untuk melihat dan membandingkan keupayaan serta prestasi setiap model dalam membuat ramalan harga bagi tempoh jangka pendek dan panjang. Prestasi ramalan juga akan dinilai berdasarkan kepada pemilihan fitur input yang menggunakan pendekatan ramalan univariat dan multivariat bagi menentukan teknik yang mampu memberikan prestasi terbaik kepada kesemua model ramalan yang dibangunkan. Ia juga akan menjadi penentu kepada teknik yang akan digunakan bagi uji kaji ramalan untuk tempoh jangka panjang.

## 1.5 METODOLOGI KAJIAN

Pelaksanaan kajian ini dibuat dengan menggunakan pendekatan kuantitatif dan mengambil kaedah CRISP-DM (*Cross-Industry Standard Process for Data Mining*) sebagai metodologi kajian bagi menguji dan mengesahkan hipotesis yang telah dinyatakan di dalam pernyataan masalah sebelum ini. Metodologi CRISP-DM ini adalah satu model proses yang sering digunakan dalam kajian dan tugas perlombongan data. Bagi membolehkan uji kaji dilaksanakan, data sekunder yang dimuat turun dari laman sesawang Yahoo Finance akan digunakan bagi membangun dan menguji model pembelajaran mendalam yang dicadangkan. Secara umumnya pelaksanaan keseluruhan kajian di bawah metodologi CRISP-DM ini merangkumi enam fasa utama iaitu Pemahaman Bisnes, Pemahaman Data, Penyediaan Data, Pemodelan, Penilaian dan Pelaksanaan. Keseluruhan proses di dalam model ini adalah satu kitar hayat dan pelaksanaan dan turutan setiap fasa tidak statik. Ini bermakna ia boleh kembali kepada fasa sebelumnya sekiranya terdapat keperluan untuk penilaian dan pelaksanaan semula fasa-fasa sebelumnya.

Fasa Pemahaman Bisnes dan Pemahaman Data merupakan fasa awal bagi memahami dan menghimpunkan maklumat serta mengenal pasti jurang berkaitan topik dan permasalahan kajian yang akan dilaksanakan. Objektif kajian dan pernyataan masalah juga akan dikenal pasti dan ditentukan. Di peringkat ini juga, data yang berkaitan dan akan digunakan untuk uji kaji akan dikumpul dan difahami strukturnya. Fasa Penyediaan Data pula merujuk kepada pemprosesan data bagi menyediakan data untuk tujuan pemodelan. Fasa pemodelan pula akan membangunkan model-model yang dicadangkan menggunakan algoritma yang telah dipilih menggunakan data yang telah diproses tadi. Setelah itu, di dalam Fasa Penilaian, model-model yang telah dibangunkan akan diuji, dinilai dan dianalisa prestasi dan kemampuannya dalam membuat pengelasan dan meramal data yang tidak pernah dilihat oleh kesemua model. Perbandingan juga akan dibuat bagi menentukan model yang mampu memberikan prestasi terbaik ke atas permasalahan kajian ini. Fasa terakhir adalah fasa pelaksanaan dan implementasi di mana model akhir yang telah dinilai boleh digunakan sebagai penyelesaian kepada permasalahan kajian yang telah dinyatakan.

## **1.6 KEPENTINGAN KAJIAN**

Seperti yang telah dijelaskan sebelum ini, pelaburan di dalam pasaran saham adalah urusan niaga yang berisiko tinggi sekiranya tidak dilaksanakan dengan analisis yang terperinci dan keputusan yang tepat. Melalui analisis dan penilaian yang efisien, pelabur dapat membuat keputusan dan pelaburan di dalam pasaran saham yang rendah risikonya serta mampu memberikan pulangan dan keuntungan kepada mereka.

Oleh itu, hasil dapatan dan penemuan daripada kajian ini akan dapat membantu para pelabur khususnya bagi konteks pelaburan saham di Malaysia menggunakannya sebagai rujukan dan analisis prediktif dalam membuat keputusan pelaburan yang tepat. Ini secara tidak langsung dapat mengurangkan risiko kerugian kepada pelabur dan seterusnya membantu meningkatkan kadar keuntungan pelaburan. Apabila hasil dapatan kajian ini boleh membantu pelabur dalam tindakan mereka, ia mampu menarik lebih ramai individu terlibat dalam bidang pelaburan saham dan seterusnya merencanakan lagi urusan niaga pasaran saham di Malaysia. Keyakinan pelabur ke atas pasaran saham di Malaysia akan meningkatkan lagi perkembangan ekonominya. Ia sudah tentu serba sedikit dapat memberi kesan secara tidak langsung dalam tempoh jangka pendek dan panjang ke atas perkembangan dan pengukuhan ekonomi negara.

Selain itu, kajian ini juga dapat memberi nilai tambah dan sumbangan dari sudut ilmiah di dalam bidang pembelajaran mesin khususnya dalam skop pembelajaran mendalam. Ini juga dapat menjadi sebahagian daripada lonjakan kepada lebih banyak pelaksanaan kajian di Malaysia dalam skop pembelajaran mendalam bagi ramalan pasaran saham dan juga bidang kewangan secara amnya.

## **1.7 ORGANISASI TESIS**

Secara keseluruhannya, tesis ini merangkumi enam (6) bab utama yang dapat diringkaskan seperti berikut:

Bab 1: Bab ini menerangkan secara ringkas pelaksanaan kajian ini yang meliputi penerangan berkaitan latar belakang, pernyataan masalah, objektif, skop dan metodologi kajian.

Bab 2: Bab ini mengupas dan membincangkan lebih lanjut tentang tinjauan, bahan-bahan bacaan dan sorotan kajian-kajian lepas yang berkait dengan permasalahan, metodologi dan pelaksanaan kajian ini.

Bab 3: Bab ini menerangkan secara terperinci metodologi dan pendekatan kajian yang dilaksanakan. Ia meliputi kaedah dan pelaksanaan dari peringkat pengumpulan data, pemprosesan dan penyediaan data hingga ke peringkat pembangunan model-model pembelajaran mendalam dan penilaian ke atas prestasinya.

Bab 4: Bab ini menganalisis dan menghuraikan hasil dapatan daripada uji kaji yang dilaksanakan. Huraian dan perbandingan prestasi setiap model yang dibangunkan dibuat bagi mendapatkan model ramalan yang terbaik.

Pusat Sumber  
FTSM

## **BAB II**

### **KAJIAN KEPUSTAKAAN**

#### **2.1 PENGENALAN**

Bab ini akan membincangkan dan menghuraikan mengenai pelbagai kajian dan penyelidikan yang telah dilakukan dalam melaksanakan ramalan ke atas data siri masa pasaran saham. Fokus utama huraian akan diberikan ke atas kajian yang menggunakan teknik pembelajaran mendalam ke atas data siri masa tersebut. Dapatan daripada semua kajian ini akan digunakan bagi mengenal pasti jurang penyelidikan yang berkaitan dan teknik serta kaedah yang bersesuaian untuk digunakan dalam pelaksanaan tajuk kajian ini. Secara umumnya, bab ini akan mengulas mengenai asas dan latar belakang analisis pasaran saham berdasarkan kepada pelbagai perbincangan di dalam yang kajian lepas. Selain itu, pelaksanaan ramalan pasaran saham juga akan dibincangkan terutama dari segi teknik dan pendekatan yang telah digunakan di dalam kajian-kajian yang lepas. Penggunaan teknik pembelajaran mendalam akan turut dibincangkan sebagai fokus utama berasaskan kepada variasi teknik dan kaedah yang dilakukan di dalam kajian-kajian yang berkaitan. Rumusan dari semua ulasan kepustakaan ini akan dibuat sebagai panduan dan rujukan bagi menentukan hala tuju dan pelaksanaan yang akan dilaksanakan sebagai metodologi dan analisis kajian kelak.

#### **2.2 LATAR BELAKANG ANALISIS PASARAN SAHAM**

Analisis pasaran saham merujuk kepada penilaian yang dilakukan oleh pelabur bagi membolehkan mereka mengenal pasti peluang yang wujud sebelum melakukan pelaburan ke atas pasaran saham tersebut. Pelbagai teori dan konsep berkaitan pasaran kewangan telah wujud bagi menerangkan atau memahami sifat-sifat dan kecenderungan pergerakan pasaran saham sama ada ianya boleh dianalisis dan diramal (Shah et al. 2019). Teori yang paling utama dan popular berkaitan dengan pasaran kewangan



adalah teori Hipotesis Pasaran Cepak (*Efficient Market Hypothesis* – EMH) yang telah dinyatakan oleh Fama (1970). Teori kontroversi ini menyatakan tentang kecekapan pasaran saham yang menyebabkan harga saham mencerminkan semua maklumat berkaitan dengan pasaran tersebut. Menurut teori ini juga, saham sentiasa diniagakan pada nilai pasaran saksama sehingga sesuatu merubahnya. Oleh itu, adalah mustahil bagi pelabur menentukan ketepatan untuk membeli saham pada harga yang rendah atau menjualnya pada harga yang tinggi bagi mendapatkan pulangan. Ini bermakna pelabur tidak dapat menjangka atau membuat ramalan yang tepat mengenai kesesuaian dan ketepatan tindakan mereka dalam membuat pelaburan kecuali mereka perlu memilih pelaburan yang berisiko tinggi bagi mendapatkan pulangan yang lebih tinggi.

EMH dikategorikan kepada tiga jenis keadaan iaitu (i) kecekapan lemah di mana pergerakan harga saham adalah secara rawak dan perubahan harga tersebut bersifat tidak bersandar antara satu sama lain, oleh itu adalah mustahil untuk mengalahkan pasaran saham dan mendapat keuntungan melalui analisis teknikal; (ii) kecekapan sederhana di mana harga saham berubah dengan pantas mengikut pasaran dan maklumat awam seperti data kewangan syarikat, ekonomi atau politik, oleh itu adalah mustahil untuk pelabur mendapat keuntungan melalui analisis fundamental; dan (iii) kecekapan kuat di mana harga saham dipengaruhi oleh pasaran, maklumat awam dan persendirian di mana tiada pelabur mampu mendapat keuntungan secara konsisten walaupun menggunakan maklumat persendirian yang menjadi maklumat awam (Bustos & Pomares-Quimbaya 2020; Shah et al. 2019; Marwala 2014; Yiing & Thim 2015; Fister et al. 2019).

Sebuah lagi teori popular yang turut menyokong teori EMH ini adalah Hipotesis Jalan Rawak (*Random Walk*) yang dinyatakan oleh Fama (1995). Ia menyatakan bahawa perubahan harga pasaran saham adalah bersifat bebas dan tidak bergantung kepada faktor lain. Ini bermakna harga pasaran yang lepas tidak mempunyai sebarang hubungan atau maklumat dengan perubahan harga terkini. Oleh itu, sejarah perubahan dan nilai harga yang lepas tidak boleh digunakan bagi membuat analisa berkaitan jangkaan harga pasaran saham terkini. Selain itu, teori ini turut menganggap bahawa harga pasaran saham mempunyai sifat turun naik yang luar biasa dan tidak boleh

dijangka. Ini kerana apa sahaja maklumat baharu yang ada akan terus memberi kesan kepada harga saham dan tidak dapat dijangka.

Berdasarkan kepada kedua-dua teori asas ini, ramalan ke atas pergerakan dan harga saham dilihat agak mustahil untuk dilaksanakan. Walaupun begitu, kedua-dua teori ini merupakan teori yang menjadi kontroversi dan sering dipertikaikan oleh pelbagai pihak khususnya pada masa kini. Dengan perkembangan teknologi dan pengetahuan, pelbagai kajian telah dibuat dan mampu membuktikan bahawa pergerakan dan nilai pasaran kewangan termasuk pasaran saham boleh diramal sekiranya dibina dan direka bentuk dengan algoritma yang bersesuaian dan optimum (Mehtab & Sen 2020; Mehtab et al. 2021). Selain itu, pengkritik kedua-dua teori ini turut memberikan senario sebenar yang menidakkan kebenaran teori ini di mana jutawan Amerika iaitu Warren Buffett telah lama menggunakan analisis pasaran kewangan dan mampu mengatasinya secara konsisten untuk tempoh yang lama (Hu et al. 2015). Ini seterusnya menafikan ketepatan kedua-dua teori tersebut khususnya dalam meramal pergerakan dan harga saham di pasaran. Disebabkan itu, para penyelidik amnya sentiasa membuat kajian dan penambahbaikan ke atas penyelesaian permasalahan ini khususnya bagi memahami dan mendapatkan trend pergerakan pasaran saham serta kaedah untuk meramal harga saham dengan lebih cepat dan berkesan.

Secara asasnya, dua kaedah konvensional dan masih digunakan sehingga kini untuk membuat analisis pasaran saham adalah analisis fundamental dan analisis teknikal (Manjunath et al. 2021; Bustos & Pomares-Quimbaya 2020; Shah et al. 2019). Kedua-dua kaedah ini biasa digunakan oleh pelabur untuk membantu mereka membuat keputusan dalam memilih pelaburan yang dijangka mampu memberikan pulangan yang tinggi dan berisiko rendah. Analisis fundamental adalah penilaian faktor-faktor yang mempengaruhi dan memberi impak kepada keuntungan syarikat yang berkenaan. Analisis teknikal pula merujuk kepada analisa data harga saham dan nilai dagangan yang lepas bagi menjangka pergerakan harga saham pada masa hadapan. Kedua-dua analisis ini akan memberikan pelabur petunjuk sama ada perlu membeli atau menjual saham yang berkaitan dan berpotensi untuk mendapat keuntungan.

Menurut Hu et al. (2015), analisis fundamental dilakukan berasaskan kepada tiga aspek penting iaitu (i) analisis makro ekonomi seperti Keluaran Dalam Negari Kasar (KDNK) di mana analisa dibuat ke atas keadaan makro ekonomi ke atas keuntungan syarikat pada masa hadapan; (ii) analisis industri di mana anggaran nilai syarikat dibuat berasaskan status semasa industri berkaitan; dan (iii) analisis syarikat di mana analisa ke atas kewangan syarikat dibuat bagi menganggar nilainya. Analisis teknikal pula dibahagikan kepada beberapa domain iaitu analisis sentimen, aliran tunai, set data harga saham, trend, momentum, jumlah dagangan dan pergerakan turun naik harga saham.

Sejak kebelakangan ini, dengan pelbagai kajian dan analisis mengenai pasaran saham yang telah dilakukan oleh para penyelidik, pakar ekonomi dan penganalisis data, semakin banyak faktor dan komponen ramalan yang telah dikenal pasti boleh digunakan dalam analisis pasaran saham. Selain daripada sejarah harga dan rekod dagangan saham yang lepas, pasaran saham juga sentiasa dipengaruhi pelbagai faktor lain seperti ekonomi, persepsi pelabur, politik dan lain-lain (H. Liu & Long 2020). Permasalahan data siri masa pasaran saham yang tidak linear dan tidak pegun juga memberikan satu cabaran baharu kepada penyelidik untuk membuat ramalan harga saham dengan tepat. Oleh itu, pembangunan satu model ramalan yang efektif dan fleksibel sudah tentu dapat memberi kesan dan impak yang positif kepada pelabur dan juga pasaran saham khususnya di Malaysia.

### **2.3 RAMALAN PASARAN SAHAM**

Ramalan pasaran saham sentiasa menjadi satu cabaran besar dalam sektor pasaran kewangan disebabkan oleh sifatnya yang sangat rumit, dinamik dan tidak linear. Ini membawa kepada satu keperluan untuk mencari kaedah dan penyelesaian secara saintifik yang mampu menyelesaikan permasalahan ini. Kajian dan penyelidikan berterusan sentiasa dibuat bagi meneroka dan mencari teknik serta kaedah yang boleh meramal pasaran saham secara lebih tepat. Walaupun teori EMH (Fama 1970) dan Hipotesis Jalan Rawak (Fama 1995) seolah-olah menafikan sebarang keupayaan teknik ramalan pasaran saham, dapatan daripada pelbagai hasil kajian terkini (Manjunath et al. 2021; Sezer et al. 2020) pula menunjukkan sebaliknya di mana ia membuktikan ramalan

pasaran saham ini boleh dibuat dan mampu mencapai ketepatan yang agak rasional. Abu-Mostafa & Atiya (1996) dalam kajiannya telah lama menafikan teori EMH dengan mengemukakan bukti-bukti seperti kewujudan trend pergerakan harga di dalam pasaran kewangan dan juga hubungan antara faktor ekonomi yang mempengaruhi pasaran kewangan. Kajian yang meluas serta penemuan pelbagai komponen yang sangat berkait dengan ramalan harga juga membuatkan ramalan saham sebenarnya boleh dilaksanakan.

Ramalan pasaran saham merupakan satu teknik yang efisien untuk digunakan oleh pelabur sebagai alat bantuan keputusan seperti yang digunakan di dalam sistem amaran awal (*Early Warning System – EWS*). Ia dapat memberi gambaran awal pergerakan harga pasaran saham kepada pelabur agar keputusan dan tindakan yang bersesuaian dapat diambil bagi memastikan pulangan keuntungan yang optimum daripada pelaburan yang telah dan akan dibuat. Secara umumnya, kajian mengenai ramalan pasaran saham terbahagi kepada dua kategori ramalan di mana ianya berasaskan kepada output yang diperlukan sama ada harga saham atau pergerakan harganya sahaja (Sezer et al. 2020). Ramalan pergerakan harga saham dianggap lebih memberi petunjuk dan panduan kepada pelabur berbanding meramal secara spesifik harga saham tersebut. Ini kerana maklumat pergerakan adalah asas utama yang akan menjadi pertimbangan pelabur dalam menentukan tindakan mereka. Disebabkan itu, kajian pergerakan harga saham lebih menjadi tumpuan sesetengah penyelidik berbanding dengan ramalan harga saham itu sendiri. Jiang (2021) meletakkan kategori ramalan harga saham sebagai permasalahan regresi manakala ramalan pergerakan turun dan naik harga saham pula sebagai permasalahan pengelasan khususnya dalam pendekatan pembelajaran mesin.

Kajian ramalan pasaran saham juga memberi fokus kepada dua jenis pilihan data iaitu ramalan saham tertentu secara spesifik atau ramalan indeks komposit pasaran saham. Saham secara spesifik merujuk kepada saham syarikat di dalam pelbagai jenis bursa manakala indeks komposit pasaran saham pula merujuk kepada indeks pasaran saham di sesebuah negara. Kebiasaannya indeks komposit ini menghimpunkan pelbagai saham syarikat yang paling kukuh dan besar di dalam sesebuah negara untuk dijadikan indikator kekuatan pasaran saham negara tersebut. Kedua-dua jenis ramalan ini juga

memberikan hasil dapatan dan keputusan yang berbeza disebabkan oleh sifat keduanya yang berlainan. Ramalan indeks komposit dilihat agak cenderung untuk menghasilkan keputusan yang lebih tepat berbanding ramalan saham disebabkan pergerakan turun naiknya yang tidak terlalu besar dan ketara pada satu satu masa.

## **2.4 TEKNIK RAMALAN PASARAN SAHAM**

Terdapat beberapa variasi kaedah atau teknik yang digunakan bagi membangunkan model ramalan pasaran saham. Mehtab & Sen (2020) membahagikan kesemua kaedah tersebut kepada tiga kategori iaitu melalui penggunaan model regresi, model ekonometrik dan pendekatan pembelajaran mesin serta pembelajaran mendalam. Model regresi dan ekonometrik tersebut sebenarnya adalah model yang mengadaptasikan pendekatan statistik dalam membangunkan model ramalan. Terdapat juga kajian lain yang memperincikan lagi kategori bagi kaedah yang digunakan dalam ramalan pasaran saham kepada beberapa pecahan yang lebih kecil. Namun secara umumnya, kesemua kaedah ini boleh dikategorikan kepada dua kategori utama yang lebih umum iaitu model statistik atau model linear dan pendekatan pembelajaran mesin termasuk pembelajaran mendalam (Cavalcante et al. 2016; J. Liu et al. 2019). Kedua-dua kaedah ini adalah teknik yang biasa digunakan dalam membuat analisis dan ramalan ke atas data siri masa bagi pasaran saham.

### **2.4.1 Pendekatan Statistik Dan Model Linear**

Penggunaan analisis statistik khususnya dalam membangunkan model ramalan pasaran saham telah memainkan peranan yang sangat penting sebelum kedatangan pelbagai algoritma dan kaedah baharu yang lebih moden. Walaupun dibanjiri dengan pelbagai teknik dan kaedah dalam ramalan pasaran saham, pendekatan kaedah statistik dan model linear serta variannya masih menjadi pilihan disebabkan keupayaan dan prestasinya yang masih relevan (Shah et al. 2019). Menurut Zhong & Enke (2017) antara kaedah statistik dalam ramalan data siri masa yang biasa digunakan adalah Autoregressive Moving Average (ARMA), Generalized Autoregressive Conditional Heteroskedastic (GARCH), Smooth Transition Autoregressive (STAR) dan Autoregressive Integrated Moving Average (ARIMA). Di antara kesemua kaedah tersebut, model ARIMA adalah teknik yang paling popular dan banyak digunakan

dalam kajian ramalan data siri masa (Yiing & Thim 2015). ARIMA juga telah menjadi satu kaedah asas dalam membuat ramalan ke atas data siri masa dan masih digunakan sehingga kini dalam pelbagai kajian termasuk ramalan ke atas pasaran saham (Siami-Namini et al. 2019).

Penggunaan ARIMA yang turut dikenali dengan nama metodologi Box-Jenkins sebagai model ramalan data siri masa telah dipelopori oleh pakar matematik bernama George Box dan Gwilym Jenkins pada tahun 1970 (Adebiyi et al. 2014). Sejak itu ia menjadi kaedah yang paling menonjol dalam membuat ramalan pasaran kewangan termasuk pasaran saham. ARIMA menggunakan pendekatan statistik dalam menghuraikan data siri masa dan seterusnya membuat ramalan dengan memodelkan hubungan di antara data tersebut secara linear. Ia sering dikaitkan dengan ramalan menggunakan input univariat kerana model ini hanya memerlukan data yang lepas bagi membuat ramalan masa hadapan. Kelebihan utama model ini adalah dalam membuat ramalan jangka pendek ke atas dari siri masa.

Kajian oleh Wadi et al. (2018) untuk meramal indeks harga saham di dalam Amman Stock Exchange (ASE) Jordan membuktikan model ARIMA mampu membuat ramalan jangka pendek dengan ralat yang minimum. Data yang digunakan merangkumi tempoh dari tahun 2010 sehingga 2018. Di dalam kajian tersebut, beberapa model ARIMA dengan variasi tetapan yang berbeza telah diuji dan model ARIMA (2,1,1) menghasilkan RMSE paling rendah berbanding yang lain. Namun begitu, kajian ini tidak menyertakan satu model dasar atau kajian lain sebagai perbandingan bagi melihat prestasi sebenar model ARIMA tersebut dalam membuat ramalan.

Adebiyi et al. (2014) pula membangunkan dua model ARIMA dengan menggunakan dua set data yang berbeza untuk menguji keupayaannya dalam membuat ramalan harga saham bagi sela masa berikutnya. Ramalan jangka pendek ini menggunakan data saham Nokia dari New York Stock Exchange (NYSE) bagi tahun 1995 sehingga 2011 dan data saham Bank Zenith dari Nigeria Stock Exchange (NSE) bagi tahun 2006 sehingga 2011. Hasil uji kaji menunjukkan tetapan yang menghasilkan model terbaik mampu membuat ramalan harga saham dengan ralat atau perbezaan dari harga sebenar yang kecil. Ini menunjukkan bahawa ARIMA mampu membuat ramalan

harga saham bagi tempoh jangka pendek dengan ketepatan yang agak memuaskan. Seperti kajian sebelum ini, kajian ini juga tidak membuat perbandingan prestasi model ARIMA yang dibangunkan dengan sebarang model dasar bagi melihat kemampuan sebenarnya berbanding model lain.

Walaupun kemampuan pendekatan statistik seperti ARIMA dalam membuat ramalan pasaran saham telah dibuktikan secara teori dan kajian saintifik khususnya bagi tempoh jangka pendek, namun dalam situasi dan keadaan tertentu, kaedah ini sebenarnya mempunyai kekangan dan limitasi dalam membuat ramalan yang tepat. Ini kerana pendekatan statistik yang digunakan sentiasa menganggap data pasaran saham berhubung dan tertabur secara linear. Oleh itu, model yang dibangunkan adalah berasaskan kepada data pasaran saham yang dianggap linear sedangkan hubungan antara data tersebut sangat rumit, tidak linear dan dinamik (Cavalcante et al. 2016). Ini menyebabkan model ramalan yang dibangunkan tidak dapat memahami dan mendapatkan maklumat tersembunyi dan corak tidak linear yang terdapat dalam data siri masa pasaran saham tersebut. Selain itu, pendekatan statistik ini juga memerlukan penggunaan data siri masa yang pegun di mana sifat-sifat statistiknya seperti min dan varian tidak berbeza sepanjang tempoh masa. Ini berbeza dengan persembahan pada data pasaran saham yang tidak pegun disebabkan oleh sifatnya yang sangat dinamik dan mempunyai pergerakan turun dan naik yang tidak dijangka dan rawak. Disebabkan kekangan dan limitasi pada model-model berasaskan pendekatan statistik tersebut, kajian dan penyelidikan mula beralih kepada kaedah yang lebih moden dan efisien iaitu pembelajaran mesin khususnya pembelajaran mendalam. Melalui pendekatan ini, model ramalan yang dibangunkan mampu memahami dan merekod hubungan tidak linear di dalam data siri masa pasaran saham yang mana tidak boleh dilakukan oleh kaedah berasaskan statistik. Ia juga boleh belajar dan memahami data tersebut secara lebih kompleks bagi menghasilkan keputusan ramalan yang lebih tepat.

#### **2.4.2 Pembelajaran Mesin**

Pembelajaran mesin merupakan satu cabang teknologi yang telah berkembang dengan pesat sejak kebelakangan ini di mana ia bukan lagi menjadi topik untuk kajian dan penyelidikan tetapi telah dikomersilkan sebagai teknologi yang praktikal dalam

pelbagai bidang. Pembelajaran mesin adalah satu cabang atau bidang yang menggabungkan teknologi sains komputer dan statistik serta menjadi tunjang utama dalam kepintaran buatan dan sains data (M. I. Jordan & T. M. Mitchell 2015). Ia telah memacu kepada pembangunan pelbagai algoritma dan teori baharu selaras dengan perkembangan data yang semakin cepat dan besar pada masa kini. M. I. Jordan & T. M. Mitchell (2015) mendefinisikan pembelajaran mesin sebagai sistem komputer yang mampu belajar dan mempertingkatkan lagi pengetahuannya melalui pengalaman iaitu data. Secara mudahnya, tanpa pengaturcaraan dan diprogramkan secara rumit, sistem komputer boleh dilatih untuk mempelajari sesuatu daripada data yang dijadikan sebagai input dan seterusnya membuat keputusan berdasarkan kepada apa yang telah dipelajarinya.

Pembelajaran mesin sering dikaitkan dengan terma perlombongan data di mana teknik ini mampu menganalisa data yang berskala atau bersaiz besar dengan efisien dan seterusnya mendapatkan maklumat-maklumat yang tersembunyi, penting dan berguna (Zhong & Enke 2017). Penggunaan pembelajaran mesin dalam ramalan pasaran saham juga merupakan sebahagian daripada pendekatan perlombongan data di mana maklumat daripada data-data saham yang lepas akan digunakan untuk membuat ramalan dan merangka strategi pelaburan. Pembelajaran mesin sememangnya telah menjadi teknik yang sering dikaji dan digunakan dalam ramalan pasaran saham disebabkan potensi dan kemampuannya yang sangat efisien. Melalui pelbagai kajian dan penyelidikan mengenai pembelajaran mesin khususnya dalam permasalahan data siri masa, teknik ini terbukti mampu memahami maklumat yang tersembunyi di dalam data pasaran saham yang bersifat kompleks dan mempunyai dimensi yang tidak linear (Zulqarnain et al. 2020). Ia juga mampu memahami data pasaran saham dan mencari hubungan data saham yang lepas dengan ramalan yang akan dibuat. Ini menyebabkan teknik ini sering menjadi pilihan dalam membuat ramalan pasaran saham berbanding dengan model dan pendekatan statistik seperti yang dinyatakan sebelum ini.

Bagi membuktikan kemampuan teknik pembelajaran mesin dalam membuat ramalan pasaran saham, Hsu et al. (2016) telah membuat perbandingan keupayaan model pembelajaran mesin ANN dan Mesin Sokongan Vektor (SVM) dengan model Autoregressive (AR) yang menggunakan pendekatan statistik. Sebanyak 34 set data



indeks pasaran saham digunakan bagi membuat ramalan pergerakan indeks tersebut bagi jam dan hari berikutnya. Hasil akhir menunjukkan teknik pembelajaran mesin memberikan ketepatan ramalan yang lebih baik berbanding model AR. Kajian ini juga menyatakan bahawa kemampuan pembelajaran mesin memahami data pasaran saham yang rumit serta tidak linear menjadi asas mengapa model ini mampu memberikan prestasi ramalan yang lebih baik daripada model AR. Selain itu, kajian ini berjaya membuktikan teknik ramalan yang lebih maju menggunakan pembelajaran mesin boleh digunakan bagi meramal perubahan harga dalam pasaran kewangan yang bersesuaian.

Di dalam kajian yang dilakukan oleh Marwala (2014), model pembelajaran mesin digunakan untuk meramal indeks pasaran saham JSE All Share Index di Afrika Selatan. Model Multilayer Perceptron (MLP), SVM dan Neuro-Fuzzy digunakan dan menunjukkan prestasi yang baik dalam membuat ramalan indeks dan juga pergerakan turun dan naiknya. Sebagai perbandingan, model Autoregressive Moving Average (ARMA) yang menggunakan pendekatan statistik dijadikan sebagai model dasar untuk tujuan penilaian prestasi. Hasil analisis membuktikan model pembelajaran mesin khususnya SVM menunjukkan prestasi dan keupayaan yang lebih baik berbanding model lain termasuk ARMA.

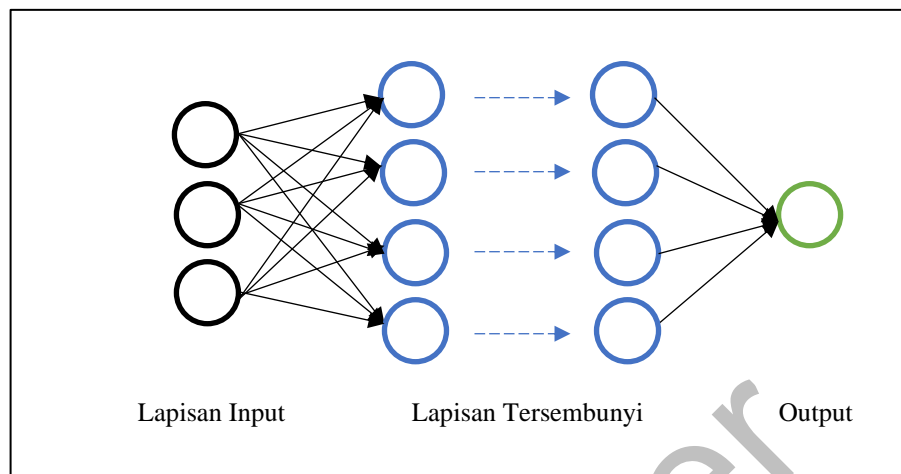
Yiing & Thim (2015) pula di dalam kajian mereka telah membuat perbandingan kemampuan model pembelajaran mesin ANN dan model ARIMA yang menggunakan pendekatan statistik dalam meramal indeks pasaran saham di FTSE Bursa Malaysia KLCI. Bagi membuat uji kaji tersebut, data dari tahun 2012 sehingga 2015 di dalam FTSE Bursa Malaysia KLCI telah digunakan untuk membangunkan model ramalan dan mengujinya. Hasil analisis menunjukkan kemampuan kedua-dua model dalam membuat ramalan untuk tempoh jangka pendek. Namun prestasi ANN sebagai model pembelajaran mesin jelas mengatasi kemampuan ARIMA dalam membuat ramalan dengan nilai RMSE paling rendah direkodkan oleh model ANN.

Teknik pembelajaran mesin terbukti telah membawa satu pembaharuan dalam pembangunan model ramalan pasaran saham khususnya dengan kekangan dan limitasi pada model berasaskan pendekatan statistik. Keupayaan teknik pembelajaran mesin memahami data yang rumit dan tidak linear menjadikannya satu kelebihan dalam

membuat ramalan pasaran saham dan mampu menghasilkan keputusan yang lebih baik. Ini terbukti dengan pelbagai kajian yang jelas menunjukkan keupayaan dan prestasinya mengatasi kaedah tradisional seperti pendekatan statistik dan model linear. Disebabkan itu, penggunaan teknik ini dalam membuat ramalan pasaran saham lebih menonjol dan menjadi pilihan utama berbanding kaedah yang menggunakan pendekatan statistik. Sejak akhir ini, teknik pembelajaran mendalam yang merupakan lanjutan daripada pembelajaran mesin pula semakin mendapat perhatian penyelidik dalam membangunkan model ramalan pasaran saham. Ia dilihat mempunyai keupayaan dan prestasi yang jauh lebih baik daripada teknik pembelajaran mesin tradisional.

## **2.5 PEMBELAJARAN MENDALAM DALAM RAMALAN PASARAN SAHAM**

Pembelajaran mendalam adalah subset kepada pembelajaran mesin dan merupakan evolusi daripada ANN di mana ia mengambil asas rekabentuk algoritma tersebut dan kebiasaannya mempunyai bilangan lapisan tersembunyi dan neuron yang banyak (Yong et al. 2017). Ia mempunyai rangkaian neural yang berlapis di antara lapisan input dan output dan direka bentuk untuk meniru bagaimana otak manusia berfikir dan belajar. Dari segi arkitektur atau rekabentuk, pembelajaran mendalam berbeza dengan pembelajaran mesin di mana ia kebiasaannya mempunyai beberapa lapisan neuron yang boleh memproses dan belajar daripada data mentah yang diberikan. Ia mempunyai struktur lapisan tersembunyi yang kompleks di mana dan mampu mempelajari corak dan maklumat tersembunyi di dalam data dengan lebih tepat (H. Liu & Long 2020). Rajah 2.1 di bawah menunjukkan contoh rekabentuk asas pembelajaran mendalam yang menggunakan beberapa lapisan rangkaian neural. Input bagi model pembelajaran mendalam tersebut akan diproses oleh beberapa lapisan tersembunyi sebelum output dikeluarkan sebagai keputusan akhir.



Rajah 2.1 Contoh struktur dan reka bentuk pembelajaran mendalam

Sumber: Nabipour et al. 2020

Walaupun rekabentuknya yang agak rumit, namun melaluinya setiap neuron akan diberikan tugas untuk menganalisis input yang mana secara tidak langsung membolehkan model pembelajaran mendalam ini menyelesaikan pelbagai jenis permasalahan rumit yang tidak boleh dibuat melalui teknik pembelajaran mesin tradisional. Pembelajaran mendalam muncul seiring dengan berkembangnya pelbagai jenis data yang semakin pesat. Perkembangan teknologi pemrosesan komputer iaitu Unit Pemrosesan Grafik (GPU) turut menyokong pembelajaran mendalam di mana ia mampu membuat pemrosesan data yang lebih laju jika dibandingkan dengan Unit Pemrosesan Pusat (CPU) (Guo et al. 2016).

Tumpuan penyelidikan dalam kajian mereka serta industri mula beralih kepada pendekatan pembelajaran mendalam berbanding pembelajaran mesin tradisional. Ini kerana ia mempunyai keupayaan untuk memproses data yang berskala besar serta boleh mempelajari hubungan tidak linear data di antara fitur input dan juga ramalan yang akan dibuat. Faktor-faktor lain seperti keupayaannya memproses dan mempelajari data raya, prestasi algoritma yang cemerlang, keupayaan memahami fitur dan maklumat tersirat, dan model yang tidak rumit untuk dibangunkan turut menyebabkan pembelajaran mendalam semakin mendapat tempat di kalangan penyelidik amnya. Dengan perkembangan dan kemampuan prestasinya, pembelajaran mendalam semakin berkembang dan digunakan dengan meluas dalam pelbagai bidang. Ini dapat dilihat melalui pelbagai kajian yang dilakukan seperti pengelasan imej (Szegedy et al. 2015),

ramalan pandemik (Zeroual et al. 2020) dan termasuk juga ramalan pasaran saham (Zulqarnain et al. 2020; Nikou et al. 2019; Hiransha et al. 2018; Yong et al. 2017).

Model pembelajaran mendalam telah banyak digunakan dalam pelbagai aplikasi berkaitan pasaran kewangan termasuk ramalan pasaran saham. Ini kerana ia mampu meningkatkan prestasi dan keupayaan ramalan jika dibandingkan dengan model pembelajaran mesin tradisional (Zulqarnain et al. 2020). Hatcher & Yu (2018) menyatakan tentang keupayaan dan prestasi model pembelajaran mendalam yang jauh lebih baik berbanding pembelajaran mesin selain mudah dikonfigurasi dan fleksibel. Jiang (2021) pula memberikan faktor seperti keupayaan pembelajaran mendalam terhadap data raya dan kebolehan memahami hubungan data yang tidak linear menyebabkannya mampu memberikan prestasi yang lebih baik berbanding model linear atau pun model pembelajaran mesin tradisional dalam pelbagai tugas termasuk ramalan pasaran saham. Ramalan pasaran kewangan sememangnya menjadi pilihan kebanyakan penyelidik sejak 40 tahun lepas. Namun dengan lonjakan prestasi melalui pembelajaran mendalam ini telah menyebabkan komuniti penyelidik mula beralih arah kepada teknik yang lebih moden ini.

Terdapat tiga jenis algoritma pembelajaran mendalam yang paling popular dan sering digunakan dalam kajian ramalan pasaran saham iaitu Multilayer Perceptron (MLP) yang merupakan subset kepada ANN, Rangkaian Neural Berlingkaran (CNN) dan Ingatan Jangka Pendek Panjang (LSTM). Setiap algoritma ini berbeza dari segi struktur dan rekabentuk serta bagaimana ia memproses dan mempelajari data.

### **2.5.1 Multilayer Perceptron (MLP)**

MLP merupakan model pembelajaran mendalam yang paling asas dan paling lama digunakan dalam pelbagai bidang. MLP adalah subset kepada ANN dan keduanya mempunyai struktur asas yang tidak berbeza dan boleh diletakkan di bawah kategori algoritma yang sama (Jiang 2021). Konsep ANN mula diperkenalkan pada tahun 1943 oleh Warren McCulloch dan Walter Pitts dan kemudiannya digunakan secara meluas untuk memodelkan pelbagai proses yang tidak linear (Nikou et al. 2019). Ia dibangunkan berasaskan konsep struktur neuron di dalam otak manusia yang mana boleh mempelajari sesuatu daripada pengalaman. MLP adalah rangkaian neural yang

mempunyai satu lapisan input, satu lapisan output dan sekurang-kurangnya satu lapisan tersembunyi. Setiap lapisan tersembunyi akan mempunyai neuron yang memproses input yang diterima berasaskan kepada pemberat dan kecenderungan yang ditetapkan. Setiap neuron pula mempunyai hubungan terus dengan kesemua neuron daripada lapisan sebelumnya. Ini akan membentuk rangkaian neural yang bersambung secara penuh di antara kesemua neuron tersebut.

MLP dan ANN merupakan antara model pembelajaran mendalam asas yang paling popular dan sering digunakan dalam kajian ramalan pasaran saham (Bustos & Pomares-Quimbaya 2020). Algoritmanya tidak terlalu rumit dan ia mampu memahami dengan lebih cekap ciri serta corak data yang mempunyai hubungan tidak linear. Kelebihan ini membuatnya sering menjadi pilihan untuk menyelesaikan pelbagai masalah regresi dan juga pengelasan. Namun begitu, MLP juga mempunyai kelemahannya tersendiri. Peningkatan jumlah dan bilangan fitur input kepada model ini akan menyebabkan peningkatan saiz parameter dan neuron di dalam rangkaian neuralnya. Ini akan memberi kesan terhadap prestasi model untuk memproses dan mempelajari data yang diberikan (Sezer et al. 2020). Bagi data siri masa seperti pasaran saham pula, elemen data berasaskan turutan masa akan dianggap tidak berkait di antara satu sama lain disebabkan struktur algoritma MLP. Ini menyebabkan maklumat daripada tempoh masa sebelumnya tidak diambil kira semasa proses mempelajari dan membuat ramalan berasaskan tempoh masa yang lain. Walaupun begitu, MLP masih menjadi teknik pembelajaran mendalam yang sesuai untuk membuat penyelesaian ke atas permasalahan ramalan pasaran saham dan mampu memberikan keputusan serta prestasi yang baik.

### **2.5.2 Ingatan Jangka Pendek Panjang (LSTM)**

LSTM adalah algoritma pembelajaran mendalam yang ditambah baik dari versi asal algoritma Rangkaian Neural Berulang (RNN). Ia mula diperkenalkan oleh Sepp Hochreiter dan Jurgen Schmidhuber pada tahun 1997 bagi mengatasi masalah kehilangan kecerunan yang berlaku pada algoritma RNN (Cao et al. 2019). Seperti RNN, LSTM juga boleh mengingati maklumat yang penting daripada input-input yang lepas dan menggunakannya untuk membuat ramalan dengan lebih tepat. Namun LSTM mempunyai keupayaan ingatan bagi tempoh jangka panjang berbanding RNN. Atas

faktor inilah maka LSTM menjadi salah satu daripada algoritma pilihan untuk data-data yang berjujukan seperti data siri masa, ucapan, teks, data kewangan, audio dan video. Struktur LSTM berbeza dari RNN di mana RNN mempunyai rangkaian neural dengan kitaran manakala LSTM mempunyai blok memori atau sel yang menggantikan struktur rangkaian neural tersebut. Setiap blok atau sel pula mempunyai 3 pintu iaitu pintu lupa, pintu input dan pintu output. Struktur rekabentuk ini membolehkan LSTM mengawal aliran maklumat dengan menentukan keperluan untuk membuang maklumat yang lepas atau menyimpannya. Ini membolehkan LSTM belajar dan memahami data yang mempunyai kebergantungan dan hubungan di antara satu sama lain seperti data pasaran saham (Livieris et al. 2020).

Konsep asas blok memori atau sel dan juga semua pintu-pintu di dalamnya adalah untuk menyimpan maklumat lepas yang telah dipelajari oleh LSTM. Pintu lupa berfungsi untuk memilih dan menentukan maklumat mana yang perlu dibuang daripada ingatan LSTM. Pintu input pula merupakan lokasi di mana maklumat baharu diterima dan pintu output pula merujuk kepada output yang dijana oleh LSTM. Algoritma LSTM sesuai digunakan bagi membuat ramalan kerana data harga saham mempunyai kaitan dan hubungan di antara satu sama lain. Melalui keupayaan ingatan jangka panjang dan pemilihan maklumat yang perlu disimpan, LSTM merupakan algoritma yang menjadi pilihan dalam pelbagai kajian melibatkan data siri masa khususnya dalam permasalahan ramalan harga saham. Namun LSTM juga mempunyai kelemahan terutamanya dari segi tempoh masa yang lama untuk melatih model terutamanya jika melibatkan jumlah dan saiz data yang besar. Ia juga memerlukan sumber ingatan komputer yang besar bagi menampung keperluan ingatan algoritma tersebut. Selain itu masalah overfitting mudah berlaku yang menyebabkan model ramalan gagal mendapat prestasi yang baik bagi data yang tidak pernah dilihat oleh model tersebut.

### **2.5.3 Rangkaian Neural Berlingkaran (CNN)**

CNN merupakan satu lagi algoritma pembelajaran mendalam yang mempunyai struktur rekabentuk yang berbeza daripada ANN dan LSTM. Ia mula diperkenalkan oleh seorang penyelidik bernama Yann LeCun pada era tahun 1980an (Hoseinzade & Haratizadeh 2019). Berbeza dengan LSTM, rekabentuk CNN memberi fokus kepada input yang diberi dan tidak mempunyai sebarang rekod atau ingatan berkaitan

maklumat lepas yang telah dipelajari (Selvin et al. 2017). CNN adalah sejenis rangkaian neural yang biasa digunakan dengan data yang mempunyai topologi seperti grid. Antara data yang sering digunakan bersama algoritma ini adalah data imej di mana data tersebut digambarkan sebagai grid dua dimensi yang mempunyai baris dan lajur. Rekabentuk CNN secara asasnya mempunyai tiga lapisan utama iaitu lapisan berlingkaran, lapisan pengumpulan dan lapisan berhubung sepenuhnya. Setiap lapisan di dalam CNN mempunyai fungsi yang berbeza sebelum maklumat dihantar kepada lapisan yang berikutnya.

Lapisan berlingkaran adalah lapisan paling utama dan lapisan asas dalam model CNN di mana di sini berlakunya proses utama ke atas input atau data yang diberikan. Lapisan ini akan mendapatkan intipati maklumat berdasarkan kepada formula pengiraan dan kernel yang telah ditetapkan di mana kernel ini akan bergerak ke setiap baris dan lajur data. Hasil akhir daripada lapisan ini akan dihantar kepada lapisan pengumpulan di mana bilangan parameter dan saiz data akan diminimakan. Pada peringkat ini sampel fitur yang paling memberi makna dan dapat mempersembahkan maklumat keseluruhan data sahaja yang akan diambil (Chung & Shin 2020). Melalui proses ini juga, hingar pada data dapat dikurangkan dan data yang lebih bersih akan diperolehi. Bagi mendapatkan fitur dengan lebih mendalam daripada jujukan data, lapisan berlingkaran dan lapisan pengumpulan boleh diulang dalam rekabentuk CNN. Lapisan terakhir pula adalah lapisan berhubung sepenuhnya di mana ia adalah sama dengan rangkaian neural di dalam ANN dan MLP. Ia mempunyai neuron yang dihubungkan kepada setiap neuron di dalam lapisan pengumpulan dan lapisan berlingkaran. Fungsi lapisan ini adalah untuk menukarkan semua fitur yang telah diperolehi daripada lapisan sebelumnya kepada output akhir sebagai keputusan yang akan dijana.

CNN juga sering digunakan di dalam ramalan pasaran saham disebabkan keupayaannya mengenal pasti fitur-fitur yang penting (Hoseinzade & Haratizadeh 2019). Ini membolehkannya memahami data harga saham yang lepas dan seterusnya membuat ramalan bagi harga pada masa hadapan. CNN juga mempunyai kelebihan ke atas data yang mempunyai hingar di mana ia boleh memilih hanya data yang penting dan mengabaikan data-data hingar. Namun kelemahan algoritma ini pula adalah ia

hanya memberi fokus dan tumpuan kepada jujukan input pada satu-satu masa sahaja dan tidak menggunakan apa yang telah dipelajari daripada input sebelumnya seperti yang dilakukan oleh LSTM. Oleh itu, secara teorinya ia akan memberi kesan kepada data siri masa yang mempunyai hubungan antara satu sama lain seperti dalam permasalahan ramalan pasaran saham.

## **2.6 PEMILIHAN DATA SAHAM DAN FITUR INPUT MODEL RAMALAN**

Pemilihan data adalah perkara paling utama yang perlu dilakukan sebelum membuat ramalan pasaran saham. Kebiasaannya kajian ramalan pasaran saham akan menggunakan data harga saham yang lepas sebagai data utama dalam membuat ramalan. Ia dipilih berdasarkan hipotesis yang menyatakan bahawa sejarah turun dan naik harga saham akan berulang. Selain itu, teori EMH menyatakan bahawa harga saham yang lepas juga turut menyimpan secara tersirat maklumat-maklumat lain yang memberi kesan kepada pergerakan harga tersebut (Jiang 2021). Oleh itu, keperluan untuk mendapatkan data-data lain bagi membuat ramalan menjadi tidak relevan dan merumitkan lagi pembangunan model ramalan. Walaupun begitu, terdapat juga kajian yang menggunakan data-data lain seperti data kewangan lain, pasaran komoditi, berita dan lain-lain. Pemilihan fitur input pula sangat memberi kesan yang ketara terhadap model ramalan yang akan dibangunkan. Fitur input berkait rapat dengan data yang dipilih dan akan digunakan bagi membuat model ramalan. Disebabkan itu, pertimbangan yang sesuai perlu diambil kira berdasarkan kepada kajian yang akan dilakukan sebelum pemilihan fitur dibuat.

### **2.6.1 Pemilihan Data**

Data pasaran saham merupakan data utama yang menjadi pilihan dalam pelbagai kajian ramalan pasaran saham berbanding data-data lain seperti data makroekonomi, teks, imej dan lain-lain. Ulasan kajian-kajian lepas yang dilakukan Jiang (2021) menunjukkan hampir 80% kajian ramalan pasaran saham dari tahun 2017 hingga 2019 menggunakan data pasaran saham sebagai data utama dalam membangunkan model ramalan. Data ini kebiasaannya mengandungi maklumat harga buka, harga tutup, harga tertinggi, harga terendah, jumlah dagangan dan lain-lain. Pembelajaran mendalam memerlukan jumlah data dalam kadar yang banyak dan besar bagi melatih modelnya yang mempunyai



algoritma yang rumit. Atas faktor ini maka data pasaran saham sering digunakan kerana ia mengandungi jumlah sampel data yang besar berbanding pilihan data-data yang lain. Selain itu, data ini mempunyai kadar hubungan yang paling tinggi dengan ramalan yang akan dibuat jika dibandingkan dengan data-data yang lain. Ini sudah tentu akan memberi kesan yang besar ke atas prestasi model dalam membuat ramalan yang lebih tepat.

Kebanyakan kajian ramalan pasaran saham lebih menumpukan kepada pemilihan dan penggunaan data saham dan indeks yang kukuh, berpengaruh dan besar terutamanya dari negara-negara seperti Amerika Syarikat, India, China, Jepun, Jerman dan Taiwan. Ini mungkin disebabkan oleh faktor pasaran saham tersebut yang telah matang, pasaran saham yang sedang meningkat naik dan status pasaran saham tersebut yang bersifat lebih global. Pasaran saham Amerika Syarikat dan indeks sahamnya menjadi tumpuan utama penyelidikan bagi kajian ramalan pasaran saham seperti S&P 500, Dow Jones Industrial, NASDAQ, NYSE dan RUSSELL. Kajian-kajian seperti H. Liu & Long (2020), Hoseinzade & Haratizadeh (2019), Sismanoglu et al. (2019), Siami-Namini et al. (2019), Cao et al. (2019) dan Di Persio & Honchar (2016) merupakan antara kajian yang menggunakan data pasaran saham dari negara tersebut.

Sementara itu pemilihan data pasaran saham Malaysia bagi kajian ramalan pasaran saham dilihat sangat terencil dan tidak menjadi pilihan. Hanya terdapat beberapa kajian berkaitannya yang dilaksanakan khususnya dalam tempoh kebelakangan ini. Antara kajian yang menggunakan data pasaran saham Malaysia adalah Al-Mashhadani et al. (2021) yang menggunakan algoritma ANN untuk meramal pergerakan harga saham Maybank. Kim Soon et al. (2018) pula menggunakan data saham CIMB bagi meramal harga saham menggunakan algoritma ANN dan NARX. Selain itu, Yiing & Thim (2015) menggunakan indeks komposit pasaran saham Malaysia sebagai data bagi membuat ramalan indeks tersebut menggunakan ANN dan membuat perbandingan prestasi model tersebut dengan model ARIMA. Yao & Poh (1995) juga menggunakan data yang sama bagi menguji kemampuan algoritma ANN dalam meramal nilai indeks tersebut. Secara keseluruhan, kesemua kajian tersebut tidak menggunakan pendekatan pembelajaran mendalam secara spesifik walaupun terdapat algoritma yang digunakan berkaitan dengan pembelajaran mendalam.

### 2.6.2 Pemilihan Fitur Input

Pemilihan fitur input untuk membangunkan model ramalan bergantung kepada pendekatan yang digunakan sama ada berasaskan kepada analisis teknikal atau analisis fundamental. Pendekatan analisis teknikal menggunakan data harga saham yang lepas atau indikator teknikal yang dijana daripada data harga saham sebagai fitur input. Analisis fundamental pula lebih menumpukan kepada fitur yang diambil daripada data luar seperti indikator ekonomi, harga saham atau aset lain, kadar faedah dan lain-lain lagi. Analisis teknikal merupakan pendekatan yang sering digunakan dalam pelbagai kajian yang melibatkan ramalan pasaran saham disebabkan ia lebih mudah untuk dibangunkan dan memberikan prestasi ramalan yang sangat baik (Bustos & Pomares-Quimbaya 2020). Pemilihan fitur input sangat memberi pengaruh yang besar terhadap prestasi dan keupayaan model ramalan yang dibangunkan. Ini kerana model ramalan akan belajar dan memahami corak data tersebut berdasarkan kepada fitur input yang diberikan kepadanya.

Berdasarkan kepada bilangan fitur input yang digunakan, model ramalan boleh dikategorikan kepada analisis univariat atau multivariat. Bagi penggunaan univariat, hanya satu pembolehubah atau fitur daripada data siri masa tersebut digunakan sebagai input. Kebiasaannya, univariat menggunakan fitur yang sama dengan sasaran ramalan yang akan dibuat. Ini bermakna sekiranya model perlu meramal harga saham pada masa hadapan, model univariat akan menggunakan harga saham yang lepas sahaja sebagai fitur input. Multivariat pula melibatkan lebih daripada satu pembolehubah atau fitur yang dijadikan sebagai input. Kebiasaannya ia melibatkan fitur daripada data pasaran saham tersebut, indikator teknikal, indikator ekonomi dan apa sahaja data luar yang boleh digunakan dan mempunyai hubungan dengan data saham tersebut (Zhong & Enke 2017). Terdapat pelbagai jenis fitur input yang digunakan bagi membuat model ramalan pasaran saham. Secara ringkasnya, berdasarkan kepada Bustos & Pomares-Quimbaya (2020) ia boleh dibahagikan kepada empat kategori iaitu harga saham, indikator teknikal, indikator makroekonomi dan data tidak berstruktur.

**a. Harga Saham**

Harga saham yang lepas merupakan fitur input yang menjadi pilihan paling utama digunakan untuk membangunkan model ramalan (Jiang 2021). Ini disebabkan kemampuan fitur ini memberikan pelbagai maklumat pasaran saham tersebut bagi meramal harga dan pergerakannya. Harga saham sebenarnya menyimpan secara tersirat sifat-sifat pasaran saham dan maklumat lain yang mempengaruhi pergerakannya (Bustos & Pomares-Quimbaya 2020). Sebagai contoh, kitaran ekonomi dunia atau sesebuah negara mempunyai pengaruh secara tidak langsung kepada pergerakan harga saham. Oleh itu, berbanding menggunakan maklumat tersebut dan pelbagai maklumat lain yang mempunyai kesan kepada pergerakan harga saham, penggunaan harga saham sebagai fitur secara logiknya sudah mengandungi apa sahaja maklumat lain yang mempengaruhinya. Kebiasaannya harga tutup saham digunakan sebagai fitur input utama kerana ia mengandungi semua maklumat bagi aktiviti saham tersebut pada hari ia didagangkan (Adebiyi et al. 2014; Marwala 2014).

**b. Indikator Teknikal**

Indikator teknikal merupakan salah satu daripada kaedah yang biasa digunakan oleh pelabur untuk melakukan analisis pasaran saham serta sebagai rujukan dan panduan dalam menentukan tindakan yang perlu dilakukan ke atas pelaburan pasaran saham mereka. Ia dijana melalui data harga saham yang lepas menggunakan pelbagai jenis kiraan dan formula bagi menghasilkan pelbagai jenis indikator teknikal (Jiang 2021). Ia juga boleh menggambarkan sifat dan trend pasaran saham yang mana menjadikannya lebih sesuai sebagai rujukan pelabur berbanding rekod harga saham semata-mata (Bustos & Pomares-Quimbaya 2020). Pemilihan indikator teknikal sebagai fitur input turut menjadi pilihan di dalam pelbagai kajian ramalan pasaran saham. Ia kebiasaannya digunakan bersama dengan harga saham sebagai fitur input di dalam analisis multivariat. Di antara contoh indikator teknikal yang biasa digunakan adalah *Simple Moving Average* (SMA) dan *Moving Average Convergence/Divergence* (MACD). Lee et al. (2021) telah membuktikan di dalam kajiannya di mana penggunaan pelbagai indikator teknikal bersama data harga saham yang lepas sebagai fitur input kepada model pembelajaran mendalam mampu menghasilkan ramalan pergerakan saham dengan kadar ketepatan yang tinggi iaitu sehingga 83%. Ia jauh berbeza dengan

penggunaan hanya satu indikator teknikal di mana purata kadar ketepatan yang dicapai hanya 75%. Ini jelas membuktikan bahawa indikator teknikal juga sesuai digunakan sebagai fitur input dalam membangunkan model ramalan pasaran saham.

#### **c. Indikator Makroekonomi**

Penggunaan indikator berkaitan ekonomi sering dikaitkan dengan pendekatan analisis fundamental dalam membuat ramalan pasaran saham. Ia digunakan untuk memahami bagaimana perubahan harga saham mempunyai kaitan dengan faktor-faktor dan perubahan luaran serta dalaman syarikat berkaitan saham tersebut (Bustos & Pomares-Quimbaya 2020). Hipotesis mengenai hubungan ini menyatakan bahawa perkembangan sesebuah syarikat yang memberi kesan langsung kepada nilai sahamnya adalah dipengaruhi oleh kedudukan ekonomi sesebuah negara tersebut. Oleh itu, berdasarkan indikator-indikator ekonomi yang berkenaan, corak pergerakan dan nilai harga saham boleh dianggar dan diramal. Di antara indikator yang sering digunakan adalah seperti kadar pertukaran matawang, KDNK, harga komoditi dan lain-lain. Selain itu, laporan kewangan syarikat juga menjadi indikator yang boleh digunakan untuk menganggar harga saham syarikat. Ini kerana kedudukan kewangan syarikat memberi kesan langsung kepada nilai sahamnya untuk tempoh masa tertentu. Walaupun indikator ekonomi turut digunakan dalam membuat ramalan pasaran saham, namun kesukaran mendapat data paling terkini untuk tempoh masa yang singkat menyebabkan kesukaran dalam mendapatkan maklumat terkini bagi membangunkan model ramalan. Ia juga kurang sesuai bagi tugas ramalan hari ke hari yang memerlukan data yang selari dengan pasaran saham terbabit.

#### **d. Data Tidak Berstruktur**

Analisis data tidak berstruktur khususnya melibatkan data teks merupakan antara topik kajian yang semakin popular pada masa kini. Walaupun ianya menjadi satu cabaran disebabkan oleh kesukaran mendapatkan maklumatnya yang tersirat serta persembahan data yang tidak berstruktur, namun perkembangan teknologi dan kepelbagaian algoritma yang ada telah menunjukkan keupayaannya yang semakin baik dan efisien. Data di dalam format teks adalah data tidak berstruktur yang sering digunakan pada masa kini dalam pelbagai kajian khususnya melibatkan pembelajaran mesin. Ramalan

pasaran saham turut menggunakan data tidak berstruktur sebagai fitur input. Data ini perlu diproses terlebih dahulu dan ditukar kepada format yang membolehkan model ramalan memahaminya. Di antara data tidak berstruktur yang sering digunakan dalam ramalan pasaran saham adalah berita berkaitan pasaran kewangan atau syarikat dan data daripada media sosial. Melalui penggunaan fitur input ini, perlombongan data teks akan dibuat bagi mendapatkan intipati berita atau sentimen serta pandangan daripada data tersebut. Kemudian data tersebut akan ditukar kepada format dan persembahan dalam bentuk nombor untuk digunakan bagi meramal pasaran saham. Namun begitu, data tidak berstruktur berbeza dengan data pasaran saham yang bersifat siri masa di mana ia berasaskan kepada situasi dan keadaan pada waktu tertentu sahaja (Bustos & Pomares-Quimbaya 2020). Oleh itu, ramalan pasaran saham menggunakan data tidak berstruktur hanya sesuai bagi membuat ramalan pada satu tempoh atau masa yang spesifik. Ini berlainan dengan konsep data siri masa yang mana ramalannya adalah bersifat secara jujukan masa data tersebut.

## **2.7 KAJIAN RAMALAN PASARAN SAHAM**

Sejak kebelakangan ini, penggunaan algoritma pembelajaran mendalam dalam kajian ramalan pasaran saham semakin mendapat tempat di kalangan penyelidik disebabkan oleh kemampuannya untuk memahami corak data tidak linear data pasaran saham secara lebih efisien. Ini terbukti dengan kewujudan pelbagai kajian berkaitannya yang merangkumi pelbagai jenis data saham dan juga algoritma pembelajaran mendalam. Tiga algoritma pembelajaran mendalam yang sering menjadi pilihan adalah ANN/MLP, CNN dan juga LSTM dengan fokus kepada ramalan harga saham dan pergerakan harga tersebut bagi tempoh jangka pendek. LSTM sering menjadi pilihan utama disebabkan kemampuan algoritma ini mengingati maklumat yang lepas di mana ianya sangat sesuai dengan struktur dan corak data siri masa yang mempunyai hubungan dengan data-data yang lepas.

Mehtab et al. (2021) membuat perbandingan beberapa algoritma pembelajaran mesin dan juga pembelajaran mendalam untuk melihat keupayaan pembelajaran mendalam meramal harga dan pergerakan saham. Data saham Tata Steel dan Hero Moto Corporation dari bursa saham India digunakan untuk melatih dan menguji kesemua model ramalan tersebut. Ia merangkumi data saham bagi setiap tempoh 5 minit

sepanjang tahun 2013 hingga 2014. Bagi tugas pengelasan pergerakan harga saham, algoritma yang digunakan adalah Regresi Linear (LR), K Jiran Terdekat (KNN), Pokok Keputusan (DT), Pembungkusan, Peningkatan, Hutan Rawak (RF), Rangkaian Neural Buatan (ANN) dan Mesin Sokongan Vektor (SVM). Tugas regresi untuk meramal harga saham pula menggunakan algoritma Regresi Multivariat, DTR, Pembungkusan, Peningkatan, RFR, ANN, SVM dan LSTM. Harga buka, harga tutup, harga tertinggi, harga terendah, jumlah dagangan serta 11 pembolehubah baharu digunakan sebagai pilihan fitur kepada kesemua model ramalan yang dibangunkan bagi membuat ramalan untuk tempoh masa seterusnya. Hasil uji kaji menunjukkan model ANN mengatasi kesemua model pembelajaran mesin lain bagi tugas pengelasan pergerakan harga saham dengan ketepatan yang paling tinggi. Bagi tugas regresi pula, model LSTM mengatasi kesemua model pembelajaran mesin termasuk model ANN dalam meramal harga saham di mana ia mencatat nilai RMSE paling rendah berbanding model-model yang lain. Walaupun kajian ini jelas membuktikan keupayaan pembelajaran mendalam mengatasi algoritma pembelajaran mesin yang lain, namun perbandingan di antara pelbagai algoritma pembelajaran mendalam yang berlainan tidak diberi tumpuan.

Kajian Lee et al. (2021) turut menggunakan algoritma LSTM dengan empat lapisan tersembunyi untuk meramal indeks pasaran saham Taiwan dan membuat perbandingan empat indikator teknikal bagi membuat ramalan tersebut. Ramalan tersebut merangkumi tugas pengelasan pergerakan indeks harga saham dan juga ramalan harga saham menggunakan data harian dari tahun 2017 hingga 2019. Selain data harga saham, empat indikator teknikal turut digunakan sebagai fitur input model ramalan. Sela masa selama 20 hari digunakan bagi menjadikan data yang lepas bagi tempoh tersebut sebagai turutan fitur input. Keputusan uji kaji menunjukkan indikator MACD berupaya menjadikan model LSTM mendapat ketepatan paling tinggi iaitu sebanyak 75% manakala kombinasi keempat-empat indikator bersama fitur input lain menghasilkan ketepatan sehingga 83%. Kajian ini hanya menumpukan keupayaan empat indikator teknikal yang dijadikan sebagai fitur input kepada model LSTM untuk membuat ramalan pasaran saham dan tidak membuat perbandingan dengan model dasar atau algoritma pembelajaran mendalam yang lain. Oleh itu keupayaan sebenar model yang dibangunkan tidak dapat dinilai dengan lebih tepat dan terperinci.

Mehtab et al. (2020) coba melihat kemampuan dua algoritma pembelajaran mendalam yaitu CNN dan LSTM dalam meramal harga saham Bharat Forge di bursa saham India. Data harian dari Desember 2012 hingga Januari 2015 digunakan dengan sela masa selama 5 hari dan 10 hari digunakan untuk dijadikan fitur input bagi meramal harga buka pada 5 hari berikutnya. Fitur input yang digunakan adalah harga buka, harga tutup, harga tertinggi, harga terendah dan jumlah dagangan bagi model multivariat manakala harga buka untuk model univariat. Beberapa tetapan dan variasi fitur input digunakan bagi melihat keupayaan model membuat ramalan. Sebanyak 3 model CNN dibangun dengan sela masa 5 dan 10 hari digunakan untuk model univariat manakala sela masa 10 hari pula untuk model multivariat. Untuk model LSTM pula, sebanyak 4 model dibangun di mana 2 daripadanya menggunakan pendekatan pengkodan dan penyahkodan. Hasil uji kaji menunjukkan kesemua model pembelajaran mendalam yang dibangun mampu memberikan prestasi yang sangat baik dengan nilai RMSE yang rendah. Secara umumnya, kesemua model CNN yang dibangun adalah model terbaik berbanding model LSTM dengan nilai RMSE serendah 0.0062. Walaupun kajian ini menggunakan dua algoritma pembelajaran mendalam dengan pelbagai tetapan pada data, namun ia hanya menggunakan satu jenis data saham dan tidak diuji dengan data saham berbeza bagi melihat keupayaan model-model yang dibangun membuat ramalan.

Kajian bagi meramal pasaran saham di bursa saham Tehran, Iran pula telah dibuat oleh Nabipour et al. (2020) dengan menggunakan algoritma pembelajaran mendalam ANN, RNN dan LSTM. Data bagi tempoh 10 tahun bermula dari 2009 hingga 2019 telah dipilih untuk digunakan bagi kajian ini. Beberapa model pembelajaran mesin turut dibangun bagi membuat perbandingan prestasi ramalan model pembelajaran mendalam. Dengan menggunakan 10 indikator teknikal sebagai fitur input dan sela masa selama 1, 2, 5, 10, 15, 20 dan 30 hari, keputusan uji kaji menunjukkan model LSTM mengatasi kesemua model-model yang lain pada setiap sela masa dengan nilai RMSE, MAPE dan MAE yang terendah. Kajian ini tidak menggunakan data mentah pasaran saham sebagai fitur input dan hanya menggunakan indikator teknikal yang dijana daripada data pasaran saham tersebut. Pendekatan ini dilihat tidak begitu menunjukkan keupayaan algoritma pembelajaran mendalam

memahami corak asal data saham tersebut dan hanya menggunakan indikator teknikal yang dilihat sebagai fitur baharu yang lebih mudah dipelajari oleh algoritma tersebut.

Liu & Long (2020) telah mencadangkan satu kerangka kerja bagi meramal harga tutup saham pada hari berikutnya dengan menggunakan algoritma LSTM sebagai rekabentuk utama. Ia menggunakan *Empirical Wavelet Transform* (EWT) dan *Outlier Robust Extreme Learning Machine* (ORELM) ke atas data pada fasa pemrosesan. Sebanyak tiga data berbeza iaitu data indeks saham S&P 500 dan Dow Jones Index dari Amerika Syarikat dan data saham China Minsheng Bank digunakan untuk kajian ini. Harga tutup saham dengan sela masa selama 10 hari ditetapkan sebagai fitur input untuk dijadikan model ramalan univariat. Perbandingan telah dibuat dengan beberapa model daripada kajian lepas dan hasil analisis menunjukkan model yang dicadangkan mengatasi kesemua model lain dengan MAPE, MAE dan RMSE yang terendah.

Maiti & Shetty D (2020) juga menggunakan algoritma LSTM dengan empat lapisan tersembunyi untuk meramal pergerakan dan harga saham bagi lima syarikat di dalam bursa saham Turki. Data-data tersebut dikumpul pada setiap masa 5 minit dan diambil dari tahun 2014 sehingga 2019. Harga buka, harga tutup, harga tertinggi, harga terendah, jumlah dagangan dan satu indikator teknikal dengan sela masa sebanyak 50 digunakan sebagai fitur input kepada model LSTM tersebut. Ramalan dibuat pada sela masa berikutnya dan hasil analisis menunjukkan ketepatan pengelasan pergerakan harga bagi kelima-lima saham mencapai 97.53% sehingga 98.91%. Bagi ramalan harga saham pula, nilai RMSE yang dicapai adalah serendah 0.024 sehingga 0.0048. Keputusan yang dicapai jelas menunjukkan kemampuan pembelajaran mendalam meramal pasaran saham dengan prestasi yang sangat baik. Namun kajian ini juga tidak menyertakan sebarang model dasar atau algoritma lain sebagai perbandingan bagi membuktikan keupayaan model LSTM tersebut dalam membuat ramalan.

Sismanoglu et al. (2019) menggunakan data saham daripada New York Stock Exchange (NYSE), NASDAQ dan NYSE MKT dari tahun 1968 hingga 2018 untuk menguji kemampuan pembelajaran mendalam meramal harga saham. Algoritma LSTM digunakan dengan fitur input yang digunakan diambil daripada data saham berkenaan iaitu harga buka, harga tutup, harga tertinggi, harga terendah dan jumlah dagangan.



Hasil analisis daripada uji kaji yang dijalankan menunjukkan model ramalan yang dibangunkan mempunyai keupayaan dan prestasi yang baik dalam meramal harga saham di mana nilai RMSE yang diperolehi adalah 0.04. Tiada sebarang perbandingan keputusan yang diperolehi dengan sebarang model lain atau kajian lepas bagi membuktikan keupayaan model yang dibangunkan untuk meramal harga saham tersebut. Selain itu, nilai RMSE yang dicatatkan masih boleh dikurangkan dengan konfigurasi dan tetapan pada rekabentuk model serta parameter berkaitannya.

Kajian oleh Fister et al. (2019) pula mencadangkan model pembelajaran mendalam yang menggunakan algoritma LSTM bagi menghasilkan satu model jual beli saham secara automatik. Ia dibangunkan berasaskan kepada tugas pengelasan pergerakan harga saham BMW di Jerman secara harian. Perbandingan dibuat dengan beberapa kaedah berasaskan syarat, indikator teknikal dan model pembelajaran mesin untuk melihat keupayaan LSTM meramal dan seterusnya membuat simulasi jual beli saham secara harian berdasarkan kepada ramalan yang dibuat. Fitur input yang digunakan adalah data saham tersebut yang mengandungi variasi harga saham serta pelbagai jenis indikator teknikal untuk meramal pergerakan pada hari pertama hingga kelima yang berikutnya. Keputusan uji kaji menunjukkan model LSTM yang dibangunkan mengatasi kesemua kaedah dan pendekatan yang ada termasuk model pembelajaran mesin. Simulasi jual beli saham berasaskan kepada model LSTM tersebut juga menunjukkan ia mampu memberi keuntungan kepada pelabur.

Bagi menambahbaik model ramalan saham yang menggunakan algoritma LSTM, Cao et al. (2019) mencadangkan algoritma *Complete Ensemble Empirical Mode Decomposition with Adaptive Noise* (CEEMDAN) untuk digunakan bersama dengan LSTM. Sebanyak empat indeks komposit saham dari negara yang berlainan digunakan sebagai data bagi kajian ini iaitu S&P 500 dari Amerika Syarikat, Hang Seng Index (HSI) dari Hong Kong, Deutscher Aktien Index (DAX) dari Jerman dan Shanghai Stock Exchange Composite Index (SSE) dari China. Kesemua data ini diambil dari tahun 2007 hingga 2017 di mana hanya harga tutup indeks saham tersebut digunakan sebagai fitur input untuk meramal harga tutup pada hari berikutnya. Prestasi model CEEMDAN-LSTM dibandingkan dengan model LSTM, SVM, CEEMDAN-SVM dan CEEMDAN-MLP bagi melihat keupayaannya untuk meramal keempat-empat indeks tersebut. Hasil

analisis menunjukkan model CEEMDAN-LSTM mengatasi kesemua model lain dengan nilai RMSE, MAE dan MAPE terendah berbanding yang lain.

Chung & Shin (2020) pula memberi fokus kepada algoritma CNN dalam kajian mereka untuk meramal pergerakan Indeks Komposit Korea (KOSPI). Data yang digunakan diambil dari tahun 2000 hingga 2016 dan sebanyak 7 indikator teknikal dijana daripadanya untuk dijadikan fitur input. Algoritma genetik (GA) digunakan sebagai kaedah untuk menambahbaik secara sistematik semua parameter di dalam model CNN. Sebagai perbandingan, model ANN dan CNN tanpa GA dibangunkan untuk melihat keupayaan model GA-CNN meramal pergerakan KOSPI. Hasil uji kaji menunjukkan model CNN jauh mengatasi model ANN di mana CNN mendapat ketepatan sehingga 70.16% manakala ANN hanya sekadar 58.62%. Dengan penambahan kaedah GA ke atas model CNN, ketepatan ramalan terus meningkat sehingga 73.74% dan jauh lebih baik daripada model asal CNN. Seperti kajian yang lain, kajian ini menggunakan indikator teknikal dan tidak menggunakan data indeks KOSPI sebagai fitur input. Walaupun keputusan ramalan yang diperolehi adalah baik, namun keupayaan CNN memahami data mentah indeks KOSPI tidak dapat dibuktikan dengan jelas kerana indikator teknikal merupakan fitur yang telah diproses daripada data mentah yang asal.

Kajian oleh Zulqarnain et al. (2020) menggunakan algoritma CNN bersama dengan Unit Berulang Berpagar (GRU) bagi meramal pergerakan indeks komposit saham HSI, S&P 500 dan DAX pada hari berikutnya. Data yang digunakan adalah dari tahun 2008 hingga 2016 di mana sela masa selama 250 hari digunakan untuk tetapan fitur input. Model GRU, DNN dan CNN dijadikan sebagai model dasar untuk tujuan perbandingan prestasi model GRU-CNN yang dicadangkan. Hasil uji kaji menunjukkan kemampuan model GRU-CNN meramal pergerakan ketiga-tiga indeks dengan ketepatan melebihi 50% di mana HIS pada kadar 56.2%, DAX pada kadar 56.1% dan S&P 500 pada kadar 56.3%. Ia lebih baik dan mengatasi kesemua model dasar yang dijadikan perbandingan untuk ketiga-tiga indeks saham. Walau bagaimanapun, kajian ini tidak menyatakan dengan jelas fitur input yang digunakan bagi membangunkan model ramalan. Selain itu, perbandingan prestasi mungkin juga boleh dibuat dengan kajian-kajian lain yang menggunakan algoritma berbeza memandangkan data yang

digunakan di dalam kajian ini adalah data yang sering digunakan dalam kebanyakan kajian ramalan pasaran saham.

Mehtab & Sen (2020b) pula turut menggunakan algoritma CNN untuk meramal indeks pasaran saham NIFTY 50 di India menggunakan fitur input multivariat. Data dari tahun 2015 hingga 2019 digunakan dengan kesemua data tersebut bersama 9 indikator tambahan digunakan sebagai fitur input. Sela masa selama 5 hari digunakan untuk meramal indeks tersebut pada hari ke 5 dan ke 10 berikutnya. Untuk perbandingan prestasi, model Regresi Multivariat, DTR, Pembungkusan, Peningkatan, RFR, ANN dan SVM digunakan untuk melihat keupayaan CNN dalam membuat ramalan. Hasil uji kaji menunjukkan model CNN yang menggunakan fitur input multivariat jauh mengatasi kesemua model pembelajaran mesin lain dengan nilai RMSE terendah. Kajian ini memberi tumpuan kepada keupayaan CNN bersama fitur multivariat dalam membuat ramalan berbanding pendekatan pembelajaran mesin. Walaupun jelas CNN mempunyai keupayaan paling baik berbanding yang lain, namun tiada perbandingan dengan algoritma pembelajaran mendalam yang lain dibuat di dalam kajian ini.

Perbandingan beberapa model pembelajaran mendalam dan juga model ARIMA telah dibuat oleh Hiransha et al. (2018). Di dalam kajian ini, data yang digunakan untuk melatih dan menguji model adalah data yang berbeza. Data saham dari bursa saham India (NSE) digunakan untuk melatih model manakala tiga data saham dari NSE dan dua data saham dari Bursa Saham New York (NYSE) digunakan untuk menguji kesemua model ramalan yang dibangunkan. Harga tutup saham digunakan sebagai fitur input dengan sela masa selama 200 hari dan ramalan pula dibuat pada hari yang ke-10 berikutnya. Empat model pembelajaran mendalam iaitu MLP, RNN, LSTM dan CNN dibangunkan manakala satu model ARIMA digunakan sebagai perbandingan prestasi model-model tersebut. Hasil uji kaji menunjukkan kesemua model pembelajaran mendalam menunjukkan prestasi yang jauh lebih baik berbanding model statistik ARIMA dengan nilai MAPE yang rendah. Model CNN pula secara keseluruhannya adalah model yang menunjukkan prestasi paling baik bagi meramal kelima-lima harga saham dengan nilai MAPE terendah.

Yong et al. (2017) memilih MLP sebagai algoritma di dalam kajian mereka untuk meramal pergerakan indeks saham FTSE STI, Singapura dan menggunakannya sebagai tindakan di dalam sistem jual beli saham. Data dari tahun 2010 hingga 2017 digunakan dengan sela masa selama 10 hari dijadikan sebagai tetapan fitur input. Fitur input terdiri daripada harga buka, harga tutup, harga tertinggi dan harga terendah dengan ramalan dibuat pada hari pertama hingga kelima berikutnya. MLP dibangunkan dengan tiga lapisan tersembunyi dan setiap lapisan tersebut mempunyai 10 neuron. Hasil uji kaji menunjukkan sistem jual beli tersebut mampu memberikan keuntungan melalui ramalan pergerakan saham yang dibuat oleh algoritma MLP. Melalui kajian ini juga, didapati nilai RMSE akan semakin meningkat apabila tempoh masa ramalan semakin besar. Kajian ini hanya memberi tumpuan kepada satu algoritma dan satu jenis data sahaja. Oleh itu keupayaan model ramalan yang dibangunkan tidak dibandingkan dengan sebarang model atau kajian lain bagi melihat prestasi sebenarnya.

Nikou et al. (2019) pula telah membuat perbandingan model ramalan yang menggunakan pembelajaran mendalam dengan beberapa model pembelajaran mesin untuk meramal indeks pasaran saham MSCI di United Kingdom. Data diambil dari tahun 2015 hingga 2018 dengan sela masa selama 10 hari digunakan sebagai tetapan fitur input. Algoritma LSTM digunakan sebagai model pembelajaran mendalam dengan MLP, SVR dan RFR digunakan sebagai perbandingan untuk melihat keupayaan dan prestasi LSTM meramal indeks tersebut. Analisis yang dibuat menunjukkan LSTM merupakan model yang lebih baik berbanding yang lain dalam meramal harga tutup indeks pasaran tersebut. Dapatan daripada kajian ini juga jelas menunjukkan pembelajaran mendalam adalah kaedah yang lebih baik untuk meramal pasaran saham berbanding kaedah lain dan ianya sesuai digunakan sebagai rujukan oleh pelabur dan penganalisa pasaran.

Sementara itu, kajian melibatkan ramalan pasaran saham di Malaysia telah dibuat oleh Al-Mashhadani et al. (2021) di mana data saham Maybank dari tahun 2016 hingga 2017 telah digunakan. Beberapa set fitur input digunakan melibatkan harga tutup, beza harga tertinggi dan terendah dan jumlah dagangan dengan sela masa selama 5 hari digunakan bagi tetapan fitur input. Ia kemudiannya digunakan pada algoritma ANN untuk meramal pergerakan harga saham pada hari berikutnya. Hasil uji kaji

menunjukkan model ramalan yang dibangunkan mampu membuat ramalan dengan lebih tepat menggunakan hanya harga tutup sebagai fitur input. Kim Soon et al. (2018) pula turut menggunakan algoritma ANN untuk meramal harga saham CIMB pada hari berikutnya. Data dari tahun 2000 hingga 2015 digunakan dengan sela masa hanya selama 5 hari untuk tetapan fitur input. Fitur input yang digunakan pula adalah harga tutup, indeks KLCI, kadar faedah dan kadar tukaran matawang. Perbandingan dibuat dengan model *Nonlinear Autoregressive Network With Exogenous Input* (NARX) dan dapatan kajian menunjukkan model ANN mampu meramal dengan lebih baik di mana ia mendapat nilai MSE paling rendah berbanding NARX. Yiing & Thim (2015) membandingkan kemampuan model ANN dengan model ARIMA untuk meramal indeks KLCI. Data dari tahun 2012 hingga 2015 digunakan dan hasil uji kaji menunjukkan ANN merupakan model ramalan yang lebih baik berbanding ARIMA di mana ia mendapat nilai RMSE yang terendah. Secara umumnya kesemua kajian melibatkan ramalan pasaran saham Malaysia ini tidak menunjukkan penggunaan pembelajaran mendalam secara khusus di dalam kajian mereka. Walaupun algoritma yang digunakan adalah antara algoritma pembelajaran mendalam, namun rekabentuk model tidak menjurus ke arah teknik dan kaedah pembelajaran mendalam untuk meramal pasaran saham berkenaan.

Secara ringkasnya, kesemua kajian beserta maklumat berkaitan data, fitur, algoritma, jenis ramalan dan prestasi kajian tersebut adalah seperti di dalam Jadual 2.1 di bawah.

Jadual 2.1 Ringkasan kajian ramalan pasaran saham

Penulis	Data	Fitur Input	Algoritma	Analisis
Mehtab et al. (2021)	Tata Steel dan Hero Moto Corp. dari NSE India	Harga saham dan 11 indikator	Pengelasan: LR, kNN, DT, Pembungkusan, Peningkatan, RF, ANN dan SVM  Regresi: Multivariate Regression, DTR, Bagging, Boosting, RFR, ANN, SVM dan LSTM	ANN mendapat ketepatan paling tinggi bagi pengelasan.  LSTM mendapat RMSE paling rendah bagi regresi.
				bersambung...

...sambungan

Lee et al. (2021)	TWSE 0050	Harga saham dan 4 indikator teknikal	LSTM	LSTM mencapai ketepatan ramalan 83% dengan 4 indikator teknikal sebagai fitur input.
Al-Mashhadani et al. (2021)	Maybank dari KLSE	Harga saham	ANN	ANN mencapai ketepatan ramalan pergerakan harga saham sehingga 98%.
Mehtab & Sen (2020b)	NIFTY 50	Harga saham dan 9 indikator	Pengelasan: LR, kNN, DT, Pembungkusan, Peningkatan, RF, ANN dan SVM  Regresi: Multivariate Regression, DTR, Bagging, Boosting, RFR, ANN, SVM dan CNN	CNN mengatasi kesemua model bagi ramalan harga saham.  RF mencatat ketepatan paling tinggi bagi pengelasan.
Zulqarnain et al. (2020)	HSI, DAX dan S&P 500		GRU-CNN	Model GRU-CNN meramal pergerakan indeks saham dengan ketepatan: HSI 56.2%, DAX 56.1% dan S&P 56.3%.
Mehtab et al. (2020)	Bharat Forge dari NSE India	Harga Saham	CNN, LSTM dan EDLSTM	Kesemua model menunjukkan keupayaan yang baik. CNN paling baik dengan nilai RMSE serendah 0.0062.
Nabipour et al. (2020)	Tehran Stock Exchange	10 Indikator Teknikal	DTR, Bagging, RFR Adaboost, Gradient Boosting, XGBoost, ANN, RNN dan LSTM	LSTM mengatasi kesemua model lain dengan nilai RMSE, MAPE dan MAE terendah.

bersambung...

...sambungan				
Chung & Shin (2020)	KOSPI	7 Indikator Teknikal	GA-CNN, ANN dan CNN	CNN memberi ketetapan yang jauh lebih baik dari ANN. Kadar ketepatan GA-CNN 73.74%, CNN 70.16% dan ANN 58.62%.
Liu & Long (2020)	S&P 500, DJI dan CMSB	Harga Saham	EWT-ORELM LSTM	LSTM yang dibangunkan mengatasi model dalam kajian yang lepas.
Maiti & Shetty D (2020)	THYAO, AKBNK, ARCLK, ASELS, GARAN dari Bursa Istanbul Turki	Harga Saham dan 1 Indikator Teknikal	LSTM	LSTM mampu mencapai ketepatan di antara 97.53% hingga 98.91% bagi tugas pengelasan manakala RMSE serendah 0.024 hingga 0.0048 dicapai bagi tugas regresi.
Sismanoglu et al. (2019)	NYSE, NASDAQ dan NYSE MKT	Harga Saham	LSTM	LSTM mampu membuat ramalan harga dengan nilai RMSE 0.04.
Fister et al. (2019)	BWM	Harga Saham dan Indikator Teknikal	LSTM	LSTM mengatasi ketepatan model pembelajaran mesin dan ia mampu memberi keuntungan apabila simulasi jual beli saham dilaksanakan.
Cao et al. (2019)	S&P 500, HSI, DAX dan SEE	Harga Saham	LSTM, SVM, EMD LSTM, CEEMDAN LSTM, CEEMDAN SVM, CEEMDAN MLP	CEEMDAN-LSTM mengatasi kesemua model lain dengan nilai RMSE, MAE dan MAPE yang terendah.

bersambung...

...sambungan				
Hiransha et al. (2018)	NSE India dan NYSE	Harga Saham	MLP, RNN, LSTM, CNN dan ARIMA	Kesemua model mengatasi ARIMA namun CNN menunjukkan prestasi terbaik berbanding lain.
Kim Soon et al. (2018)	CIMB dari KLCI	Harga Saham, Indeks Komposit, Kadar Faedah dan Kadar Tukaran Matawang	ANN dan NARX	ANN model terbaik berbanding NARX dengan nilai MSE yang terendah.
Yong et al. (2017)	FTSE STI	Harga Saham	MLP	MLP mampu memberi keuntungan apabila simulasi dibuat di ke atas sistem jual beli saham.
Yiing & Thim (2015)	KLCI	-	ANN dan ARIMA	ANN menunjukkan keupayaan lebih baik berbanding ARIMA dengan nilai RMSE terendah.
Nikou et al. (2019)	MSCI United Kingdom	-	MLP, SVR, RFR dan LSTM	LSTM mempunyai keupayaan yang lebih baik berbanding model-model yang lain.

## 2.8 KESIMPULAN

Berdasarkan kepada beberapa ulasan kajian yang telah dibuat sebelum ini, jelas pendekatan pembelajaran mendalam dilihat mampu memberikan prestasi ramalan pasaran saham yang lebih baik berbanding teknik atau kaedah yang lain. Keupayaan kaedah ini memahami hubungan data yang tidak linear serta mempelajari corak dan pergerakan pasaran saham yang lepas sangat memberi kesan kepada keupayaan dalam membuat ramalan pasaran saham. Disebabkan itu, pemilihan pembelajaran mendalam



sebagai kaedah dalam kajian ini untuk meramal pasaran saham dibuat berasaskan kepada keupayaannya yang telah dibuktikan melalui kajian-kajian lepas.

Secara keseluruhan, algoritma ANN/MLP, LSTM dan CNN merupakan tiga algoritma pembelajaran mendalam yang sering digunakan dalam membuat ramalan pasaran saham. Justeru itu, ketiga-tiga algoritma ini akan menjadi pilihan untuk digunakan di dalam pelaksanaan kajian ini. Rumusan daripada semua kajian lepas yang telah dibuat juga menunjukkan fokus kajian terhadap ramalan pasaran saham di Malaysia adalah terpencil dan tidak begitu menonjol. Selain itu, penggunaan pembelajaran mendalam di dalam kajian ramalan pasaran saham di Malaysia juga tidak begitu menjadi pilihan penyelidik berbanding data saham dari negara-negara yang lebih maju atau membangun. Oleh itu, ia menjadi satu justifikasi kukuh untuk kajian ini memberi fokus kepada ramalan pasaran saham di Malaysia dengan menggunakan pendekatan pembelajaran mendalam.

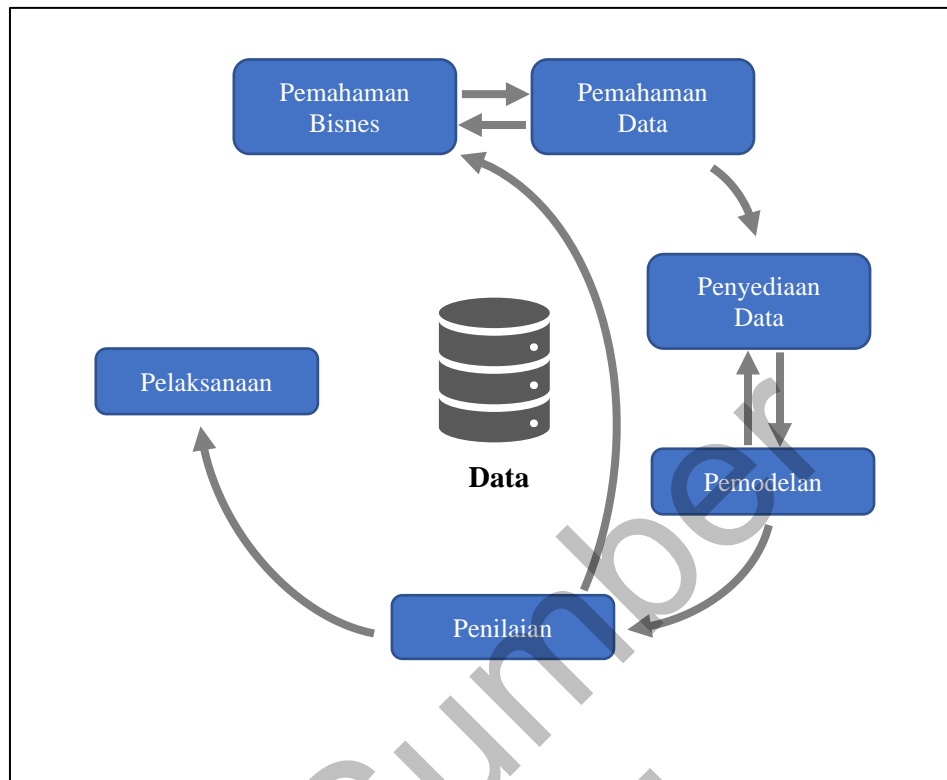
Pemilihan data saham daripada kajian lepas juga dibuat dengan mengambil kira kedudukan saham berkenaan yang kukuh di negaranya. Ia juga kebiasaannya melibatkan data saham dari syarikat yang menjadi penanda aras indeks saham di negara berkaitannya. Berasaskan kepada faktor-faktor ini, pemilihan data saham Axiata Group Berhad dan Petronas Gas Berhad bagi kajian ini adalah bersesuaian dengan kedudukannya sebagai saham yang sentiasa berada di dalam kedudukan 30 teratas di Malaysia dan menjadi penanda aras di dalam indeks komposit pasaran saham di Malaysia. Kedua-dua saham ini juga adalah melibatkan syarikat yang global dan juga relevan dengan pelaburan saham di Malaysia.

## **BAB III**

### **METODOLOGI KAJIAN**

#### **3.1 PENGENALAN**

Bab ini akan menerangkan secara lebih terperinci pendekatan dan kaedah yang digunakan dalam melaksanakan kajian. Memandangkan kajian ini melibatkan aktiviti perlombongan data, model CRISP-DM digunakan sebagai kerangka utama kepada metodologi kajian. Ia telah terbukti sebagai metodologi yang mampu membantu pelaksanaan pelbagai projek dan kajian yang berkaitan dengan perlombongan data. Model ini membahagikan pelaksanaan kajian kepada beberapa fasa utama di mana setiap fasa mempunyai aktiviti dan tugas yang perlu dilakukan. Setiap fasa juga mempunyai hubungan dengan fasa yang lain di mana ia menjadikan metodologi ini lebih dinamik dan fleksibel. Fasa-fasa tersebut adalah Fasa Pemahaman Bisnes, Fasa Pemahaman Data, Fasa Penyediaan Data, Fasa Pemodelan, Fasa Penilaian dan Fasa Pelaksanaan. Namun di dalam bab ini, Fasa Pelaksanaan tidak akan diterangkan dan hanya memberi fokus kepada fasa-fasa sebelumnya sahaja. Rajah 3.1 di bawah menunjukkan fasa-fasa di dalam model CRISP-DM dan hubungan di antaranya yang menjadi pendekatan dalam metodologi kajian ini. Aliran proses kerja ditunjukkan melalui arah anak panah di mana terdapat fasa yang boleh bergerak atau kembali kepada fasa-fasa sebelumnya. Ini membolehkan aktiviti-aktiviti yang dilaksanakan dinilai atau dikemaskini semula sekiranya terdapat keperluan atau kesilapan dalam fasa berkenaan. Fleksibiliti tersebut menjadikan metodologi ini lebih dinamik dan membolehkan pelaksanaan aktiviti di dalam setiap fasa dilakukan dengan lebih tepat dan sempurna.



Rajah 3.1 Metodologi CRISP-DM serta kitar hayat fasa di dalamnya

Pelaksanaan aktiviti di dalam keseluruhan fasa kajian ini menggunakan sepenuhnya pengaturcaraan Python. Bagi tujuan tersebut, Google Colab digunakan sebagai medium pengaturcaraan yang disediakan oleh Google secara percuma melalui capaian awan dengan versi Python 3.7.12.

### 3.2 PEMAHAMAN BISNES

Pasaran saham sebagai indikator dan gambaran kestabilan ekonomi telah berkembang dengan pesat dan meluas seiring dengan kemajuan sesebuah negara. Peningkatan keupayaan teknologi maklumat serta kemudahan capaian maklumat yang semakin pantas membuatkan pasaran saham menjadi semakin global dan berjaya menarik penglibatan lebih ramai pelabur khususnya ke atas pasaran saham yang stabil dan kukuh. Ini telah membawa kepada perkembangan pelbagai kaedah dan teknik yang digunakan oleh pelabur sebagai rujukan dan panduan untuk menentukan tindakan ke atas pelaburan mereka. Ramalan pasaran saham sudah pastinya menjadi satu keperluan utama kepada pelabur bagi memastikan tindakan yang tepat diambil dan seterusnya mampu membawa keuntungan ke atas pelaburan mereka.

Pelbagai kaedah dan teknik yang telah wujud dan digunakan untuk meramal pergerakan dan harga pasaran saham. Pembangunan model ramalan pasaran saham menggunakan pendekatan pembelajaran mendalam dan data siri masa harga saham merupakan antara kaedah yang sering menjadi pilihan disebabkan oleh keupayaannya memberikan kadar ketepatan ramalan yang lebih baik. Selain itu, pembangunan model ramalan melalui pendekatan ini tidak melibatkan proses dan kaedah yang rumit serta menjimatkan kos dan juga masa. Ini bersesuaian dengan kehendak dan keperluan pelabur yang memerlukan maklumat serta rujukan yang pantas dan tepat bagi memastikan pelaburan mereka memberikan pulangan yang sewajarnya. Seperti yang telah diulas di dalam Bab 2 sebelum ini, teori EMH (Fama 1970) dan Hipotesis Jalan Rawak (Fama 1995) telah menafikan keupayaan model ramalan berasaskan data harga saham yang lepas. Namun pelbagai kajian yang telah dilakukan membuktikan sebaliknya di mana model ramalan tersebut boleh dibangunkan dan mampu memberikan prestasi yang sangat baik.

Melalui pemahaman ke atas kesemua dapatan dan penemuan tersebut dari sudut dan perspektif bisnes itu sendiri, objektif kajian ini telah ditetapkan untuk membangunkan satu model ramalan pasaran saham berasaskan pendekatan pembelajaran mendalam sebagai rujukan dan panduan kepada pelabur dalam menentukan tindakan pelaburan mereka. Ini disebabkan oleh keperluan pelabur kepada satu mekanisme yang lebih pantas dan cepat dalam memberikan maklumat berkaitan pelaburan yang mampu memberikan keuntungan kepada mereka. Objektif dan matlamat kajian ini juga dapat memenuhi keperluan dan permintaan di dalam pelaksanaan sebenar urusan niaga saham khususnya di kalangan pelabur kelak. Dari sudut kajian pula, keupayaan pembelajaran mendalam membaca dan memahami hubungan dan corak data harga saham lepas akan dapat dibuktikan yang mana dapat dijadikan sebagai pemangkin kepada kaedah dan teknik ramalan pasaran saham di kalangan pelabur.

### **3.3 PEMAHAMAN DATA**

Sumber data yang digunakan di dalam kajian ini dimuat turun daripada laman sesawang Yahoo Finance di mana ia menyediakan capaian ke atas data saham di seluruh dunia.

Bagi tujuan kajian ini, set data yang dimuat turun daripada laman sesawang tersebut adalah data saham daripada Bursa Malaysia yang melibatkan data Axiata Group Berhad dan data Petronas Gas Berhad. Kedua-dua data ini diambil bagi tempoh lima tahun iaitu bermula daripada bulan Januari 2017 sehingga Disember 2021. Tempoh lima tahun terkini digunakan sebagai sampel data untuk kajian kerana ia dapat memberi lebih ketepatan ke atas analisis dan ramalan yang akan dilakukan kelak berbanding tempoh yang terlalu lama di mana maklumat dan corak data tersebut mungkin tidak relevan untuk ramalan terkini (Shen & Shafiq 2020). Tempoh masa yang terlalu pendek pula akan mengakibatkan algoritma pembelajaran mendalam tidak dapat mempelajari dan memahami data dengan lengkap dan sempurna. Selain itu, tempoh masa tersebut turut menjadi antara pilihan yang sering digunakan penyelidik dalam pelbagai kajian lepas mengenai ramalan pasaran saham (Jiang 2021). Data-data saham ini merupakan data siri masa disebabkan ia direkod dan dikumpul berdasarkan kepada jujukan dan sela masa yang tetap iaitu pada setiap hari.

Sebagai persediaan sebelum data-data ini diproses untuk penyediaan semasa fasa pemodelan, ia perlu dianalisis terlebih dahulu bagi memahami struktur dan rekabentuknya. Ini bagi menentukan tindakan dan teknik bersesuaian yang perlu dilakukan ke atas data tersebut semasa fasa penyediaan data nanti bagi membolehkan model ramalan dibangunkan. Untuk pemahaman umum ke atas persembahan data-data ini, atribut atau fitur data tersebut akan diperincikan bagi menerangkan jenis maklumat yang direkod melalui data-data ini. Kedua-dua data yang digunakan untuk kajian ini mempunyai struktur atribut dan fitur yang sama di mana perbezaan yang wujud hanya pada nilainya sahaja. Set data ini mempunyai tujuh fitur di mana enam daripada fitur tersebut adalah jenis nilai numerik atau berterusan manakala satu fitur lagi adalah jenis sela masa. Nilai bagi fitur Open, High, Low, Close dan Adj Close adalah melibatkan harga-harga saham berkenaan di dalam Ringgit Malaysia (RM). Fitur Date pula menyimpan tarikh harian manakala fitur Volume pula menunjukkan bilangan saham tersebut diniagakan pada hari atau tarikh berkenaan. Jadual 3.1 di bawah menunjukkan kesemua senarai fitur bagi kedua-dua set data kajian ini serta penerangan ringkas mengenainya.

Jadual 3.1 Senarai fitur bagi set data kajian dan penerangan ringkas mengenainya

<b>Nama Fitur</b>	<b>Jenis Data</b>	<b>Penerangan</b>
Date	Sela masa	Tarikh data direkod
Open	Numerik	Harga buka saham
High	Numerik	Harga tertinggi saham
Low	Numerik	Harga terendah saham
Close	Numerik	Harga tutup saham
Adj Close	Numerik	Harga tutup saham setelah ditolak bayaran lain
Volume	Numerik	Jumlah dagangan

Disebabkan data ini adalah data siri masa, maka fitur Date berperanan sebagai fitur yang merekod sela masa bagi nilai setiap fitur yang lain secara harian. Fitur tersebut menyimpan tarikh urus niaga saham tersebut diniagakan. Ini bermakna nilai fitur-fitur yang lain akan menyimpan maklumat yang berkenaan mengikut tarikh urus niaga secara harian. Fitur Open menyimpan nilai harga saham dibuka untuk urus niaga pada hari berkenaan. Fitur Close pula merujuk kepada nilai harga saham semasa urus niaga pada hari berkenaan ditutup. Fitur High dan Low merujuk kepada harga tertinggi dan harga terendah pada hari yang berkenaan sepanjang urus niaga dijalankan. Fitur Adj Close adalah harga tutup saham setelah ditolak bayaran dan caj tertentu manakala fitur Volume merujuk kepada jumlah dagangan bagi saham tersebut pada hari berkenaan. Kebiasaannya pelabur menjadikan harga tutup sebagai indikator jangkaan utama untuk menentukan tindakan di dalam urus niaga mereka sama ada perlu menjual, membeli atau menahan saham berkenaan. Ini kerana harga tutup pada setiap hari transaksi akan menentukan sama ada pelaburan mereka pada hari tersebut mampu memberikan pulangan dan keuntungan atau tidak.

### **3.3.1 Data Axiata Group Berhad**

Axiata Group Berhad yang turut dikenali dengan nama Axiata merupakan syarikat multinasional Malaysia yang terlibat di dalam industri telekomunikasi. Ia merupakan satu konsortium yang menggabungkan beberapa syarikat di bawahnya dan memberi tumpuan ke arah mengembangkan operasi mereka di benua Asia. Syarikat ini bermula sebagai sebahagian daripada Telekom Malaysia pada tahun 1992, ia kemudiannya menjadi entiti berasingan dan mula disenaraikan di dalam papan utama Bursa Malaysia pada tahun 2008. Sebagai syarikat telekomunikasi yang terkemuka di Asia dan

disenaraikan di antara 30 syarikat utama di dalam KLCI, ia pasti menjadi sasaran dan tumpuan pelabur di dalam pasaran saham Malaysia untuk dijadikan sebagai prospek pelaburan yang mampu memberi keuntungan. Data Axiata yang dimuat turun dari tahun 2017 hingga 2021 mempunyai sejumlah 1233 rekod. Ringkasan penerangan statistik secara keseluruhan data tersebut adalah seperti di dalam Jadual 3.2 di bawah.

Jadual 3.2 Ringkasan penerangan statistik data Axiata Group Berhad

	Open	High	Low	Close	Volume
<b>Bilangan</b>	1233	1233	1233	1233	1233
<b>Min</b>	4.074898	4.123085	4.026348	4.078396	4933180
<b>Sisihan Piawai</b>	0.58618	0.585127	0.59067	0.591957	4041354
<b>Minimum</b>	2.598543	2.686463	2.598543	2.598543	0
<b>25%</b>	3.692666	3.74339	3.634053	3.69	2498900
<b>50%</b>	4.05	4.099098	3.995744	4.051497	3962600
<b>75%</b>	4.530327	4.575121	4.492769	4.53819	6301200
<b>Maksimum</b>	5.296682	5.408192	5.240927	5.296682	47032900

Secara umumnya, purata harga buka, tertinggi, terendah dan tutup bagi saham Axiata ini tidak menunjukkan jurang yang terlalu besar. Sisihan piawai juga menyokong pernyataan ini di mana nilai bagi setiap fitur tersebut masih di dalam kadar yang rendah. Ini menunjukkan harga saham ini tidak mengalami kadar turun dan naik yang begitu drastik. Namun dapat dilihat beza harga minimum dan maksimum bagi kesemua fitur masih menunjukkan jurang yang agak besar di mana mampu memberikan kesan ke atas pelaburan. Rajah 3.2 di bawah menunjukkan sebahagian contoh data Axiata Group Berhad.

	A	B	C	D	E	F	G
1	Date	Open	High	Low	Close	Adj Close	Volume
2	2/1/2017	4.72	4.72	4.72	4.72	4.318914	0
3	3/1/2017	4.57	4.59	4.5	4.52	4.13591	3061100
4	4/1/2017	4.53	4.53	4.46	4.51	4.12676	4707200
5	5/1/2017	4.5	4.66	4.49	4.65	4.254863	2626100
6	6/1/2017	4.65	4.93	4.62	4.93	4.51107	11006200
7	9/1/2017	4.87	4.87	4.73	4.75	4.346365	3462900
8	#####	4.75	4.82	4.68	4.74	4.337215	3932000
9	#####	4.74	4.77	4.7	4.72	4.318914	2145700

Rajah 3.2 Contoh data mentah Axiata Group Berhad

### 3.3.2 Data Petronas Gas Berhad

Petronas Gas Berhad merupakan anak syarikat yang dimiliki sepenuhnya oleh Petronas dan terlibat di dalam industri utiliti gas dan infrastrukturnya. Ia terlibat sepenuhnya dalam aktiviti pemprosesan gas asli dan penghantarannya ke seluruh Malaysia dan Singapura. Ia disenaraikan di dalam papan utama Bursa Malaysia pada tahun 1995 dan turut diletakkan di dalam KLCI sebagai 30 syarikat utama. Atas faktor ini dan kestabilan serta kemampuan syarikat yang kukuh, pelabur pasti akan tertarik dan menjadikan saham syarikat ini sebagai sebahagian tumpuan utama mereka di dalam pelaburan pasaran saham Malaysia.

Jumlah rekod di dalam data Petronas Gas Berhad yang dimuat turun dari tahun 2017 hingga 2021 adalah sebanyak 1233. Ringkasan penerangan statistik bagi data ini adalah seperti di dalam Jadual 3.3 di bawah.

Jadual 3.3 Ringkasan penerangan statistik data Petronas Gas Berhad

	Open	High	Low	Close	Volume
<b>Bilangan</b>	1233	1233	1233	1233	1233
<b>Min</b>	15.527712	15.675962	15.365538	15.51889	879154
<b>Sisihan Piawai</b>	0.770652	0.761878	0.774404	0.771884	1036022
<b>Minimum</b>	13.232759	13.296083	12.311258	13.296083	0
<b>25%</b>	15.06821	15.237514	14.920414	15.076316	407300
<b>50%</b>	15.522783	15.671988	15.374277	15.523977	652200
<b>75%</b>	16.015854	16.159817	15.829805	16.036081	1068400
<b>Maksimum</b>	17.583499	17.583499	17.508524	17.508524	24535600

Berdasarkan ringkasan penerangan statistik ini, jelas harga saham Petronas Gas Berhad lebih tinggi berbanding Axiata di mana purata keseluruhan fitur berada pada kadar RM15.00. Purata keseluruhan fitur pula tidak terlalu jauh berbeza dan tidak menunjukkan jurang harga yang besar. Sisihan piawai bagi data ini lebih besar berbanding data Axiata di mana ia menunjukkan kadar turun dan naik harga saham ini berada dalam julat yang lebih besar. Ini dapat dilihat pada harga minimum dan maksimum untuk kesemua fitur di mana perbezaannya menunjukkan nilai yang agak besar dan memberi kesan ke atas pelaburan saham ini apabila perubahan tersebut berlaku secara mendadak. Contoh sebahagian data Petronas Gas Berhad adalah seperti di dalam Rajah 3.3 di bawah.



	A	B	C	D	E	F	G
1	Date	Open	High	Low	Close	Adj Close	Volume
2	2/1/2017	21.3	21.3	21.3	21.3	17.21179	0
3	3/1/2017	21.26	21.46	21.26	21.38	17.27643	626100
4	4/1/2017	21.48	21.56	21.38	21.48	17.35724	363000
5	5/1/2017	21.54	21.62	21.48	21.5	17.3734	1333900
6	6/1/2017	21.76	21.76	21.46	21.5	17.3734	1214000
7	9/1/2017	21.5	21.74	21.3	21.4	17.2926	1787600
8	10/1/2017	21.58	21.58	20.9	20.98	16.95321	2144900
9	11/1/2017	21	21.02	20.36	20.66	16.69463	2365000

Rajah 3.3 Contoh data mentah Petronas Gas Berhad

### 3.4 PENYEDIAAN DATA

Data mentah tidak memberi apa-apa makna kepada model ramalan sekiranya ia tidak diproses terlebih dahulu supaya algoritma yang akan digunakan kelak boleh membaca dan memahami data tersebut. Oleh itu, selepas data-data dianalisis dan difahami, ia perlu diproses dan disediakan dalam rekabentuk dan format yang boleh dibaca oleh model pembelajaran mendalam. Beberapa aktiviti pra-pemprosesan data perlu dilakukan ke atas data-data mentah melibatkan pembersihan data, pengurangan data, dan transformasi data supaya struktur akhir data tersebut tersedia dan boleh digunakan di dalam fasa pemodelan nanti.

#### 3.4.1 Pembersihan Data

Data mentah kebiasaannya perlu dibersihkan terlebih dahulu sebelum ia benar-benar mempunyai kualiti dan konsistensi pada setiap fitur atau attribut data tersebut. Pembersihan data merupakan aktiviti utama yang perlu dilakukan di peringkat awal dalam pra-pemprosesan data kerana ia boleh memperbaiki kualiti data dengan membuat pembetulan ke atas data yang tidak tepat, tidak lengkap atau tidak munasabah dengan set data sebenar. Semakan ke atas data-data menggunakan pengaturcaraan Python dibuat bagi mengenal pasti nilai pada rekod dan fitur yang hilang. Sekiranya terdapat kehilangan pada data tersebut, tindakan perlu dibuat sama ada menghapuskan rekod yang terlibat atau menggantikan nilai yang hilang dengan suatu nilai yang bersesuaian dan tidak menjejaskan kualiti asal data tersebut. Rajah 3.4 di bawah menunjukkan

keputusan semakan kehilangan nilai yang dibuat pada kedua-dua set data yang akan digunakan dalam kajian ini menggunakan pengaturcaraan Python.

<code>df_axiata.isnull().sum()</code>	<code>df_petgas.isnull().sum()</code>
Open 0	Open 0
High 0	High 0
Low 0	Low 0
Close 0	Close 0
Volume 0	Volume 0
Dividends 0	Dividends 0
Stock Splits 0	Stock Splits 0
dtype: int64	dtype: int64

Rajah 3.4 Semakan nilai yang hilang pada set data kajian

Kod aturcara ini akan menyatakan bilangan nilai yang hilang pada setiap fitur yang berlaku di dalam set data. Nilai 0 menunjukkan tiada kehilangan nilai yang berlaku pada fitur data. Disebabkan tiada kehilangan nilai berlaku pada kedua-dua data, aktiviti pembersihan data tidak perlu dilakukan ke atas set data tersebut.

### 3.4.2 Pengurangan Data

Ramalan melibatkan data siri masa secara asasnya menggunakan fitur yang sama untuk input dan juga output. Ini bermakna hanya fitur harga tutup sahaja yang digunakan sebagai fitur input disebabkan ramalan yang akan dibuat adalah untuk membuat jangkaan bagi harga tutup saham berkenaan. Analisis dan ramalan ini dikenali sebagai univariat. Namun begitu, tambahan fitur lain dari set data yang sama juga sering digunakan untuk memberi lebih maklumat kepada algoritma yang akan digunakan supaya ramalan yang dibuat akan memberi keputusan yang lebih tepat. Kaedah ini dikenali sebagai ramalan menggunakan multivariat. Bagi tujuan kajian ini, kedua-dua kaedah ini akan digunakan sebagai pilihan fitur input untuk mempelbagaikan kaedah dan teknik ramalan dan memilih pendekatan yang mampu menghasilkan keputusan dan ramalan yang paling tepat.

Harga tutup merupakan fitur yang akan digunakan sebagai input dalam ramalan menggunakan pendekatan univariat manakala harga tutup, harga buka, harga tertinggi,

harga terendah dan jumlah dagangan pula digunakan untuk multivariat. Oleh itu, daripada data mentah yang asal, pengurangan data akan dibuat dengan menghapuskan fitur atau lajur Adj Close disebabkan ia tidak akan digunakan untuk fasa pemodelan kelak. Struktur data asal akan berubah di mana hanya fitur Date, Open, High, Low, Close dan Volume sahaja yang kekal untuk digunakan nanti. Rajah 3.5 di bawah menunjukkan struktur set data selepas fitur lain dihapuskan daripada data tersebut.

	Open	High	Low	Close	Volume
Date					
2017-01-02	4.318914	4.318914	4.318914	4.318914	0
2017-01-03	4.181661	4.199961	4.117609	4.135910	3061100
2017-01-04	4.145060	4.145060	4.081008	4.126760	4707200
2017-01-05	4.117609	4.264013	4.108459	4.254863	2626100
2017-01-06	4.254863	4.511069	4.227412	4.511069	11006200

Rajah 3.5 Contoh set data selepas fitur yang tidak berkenaan dihapuskan

### 3.4.3 Transformasi Data

Transformasi data merujuk kepada proses menukar format data mentah kepada satu bentuk atau format yang berbeza bagi membolehkan algoritma pembelajaran mendalam memahami data atau menambahbaik keupayaan dan proses algoritma mempelajarinya. Proses ini menjadi kunci utama kepada penentu prestasi algoritma dan model ramalan mencapai tahap yang paling optimum dalam memahami data. Bagi mencapai tujuan tersebut, dua proses atau aktiviti transformasi data akan dilakukan iaitu proses normalisasi data dan proses menukar set data kepada format data berselia.

#### a. Normalisasi Data

Disebabkan nilai setiap fitur di dalam set data yang akan digunakan sebagai input kepada algoritma pembelajaran mendalam adalah jenis numerik, normalisasi akan dilakukan supaya ia akan diskalakan kepada julat yang lebih kecil iaitu di antara 0 hingga 1. Kebiasaannya normalisasi dilakukan supaya nilai kesemua fitur yang

mempunyai jurang yang terlalu besar dapat diskalakan kepada satu julat yang standard dan kecil. Ini bagi membolehkan algoritma pembelajaran mesin dan pembelajaran mendalam memahami data dengan lebih efisien dan cepat berbanding menggunakan nilai numerik asal. Walaupun jurang nilai di antara fitur dalam set data kajian tidak terlalu besar, namun normalisasi data akan memberi kesan yang lebih efektif ke atas algoritma pembelajaran mendalam untuk mempelajari data berkenaan. Disebabkan itu, normalisasi akan dilakukan ke atas set data kajian dengan menetapkan julat nilai setiap fitur berada di antara 0 hingga 1. Formula pengiraan untuk menukar nilai numerik fitur menggunakan kaedah normalisasi adalah seperti berikut:

$$X = \frac{x - \min}{\max - \min}$$

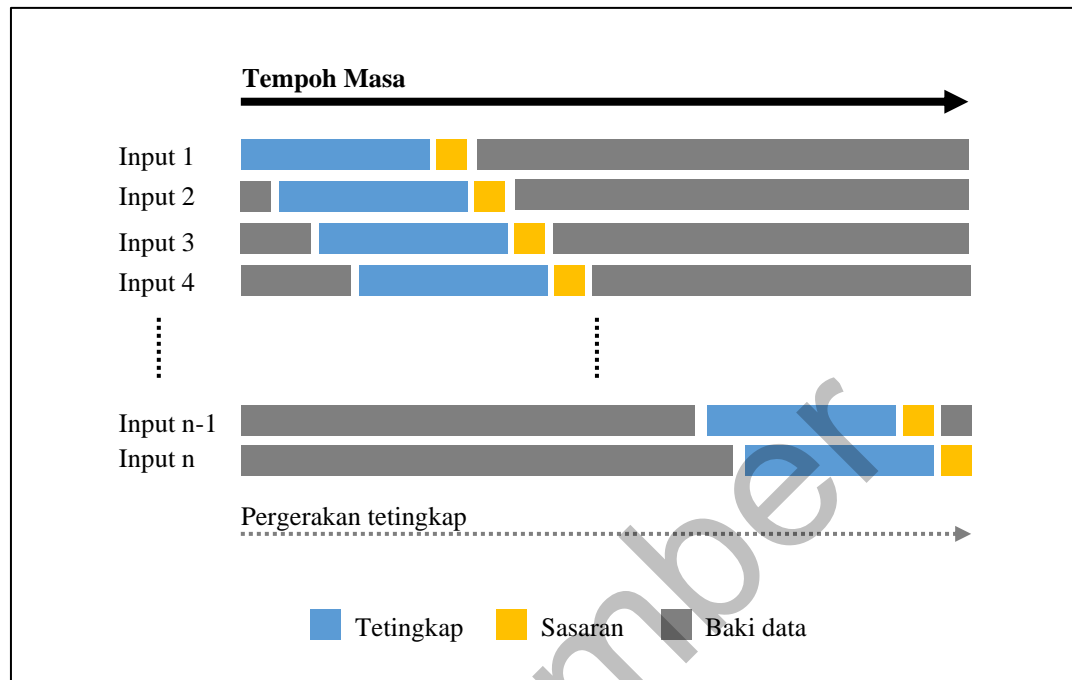
di mana X adalah nilai setelah normalisasi, x adalah nilai asal, min adalah nilai minimum dan max adalah nilai maksimum. Bagi data kajian ini, teknik normalisasi akan dilakukan ke atas semua nilai fitur termasuk fitur ramalan iaitu harga tutup.

#### **b. Pembelajaran Berselia Menggunakan Kaedah Tetingkap**

Pembelajaran berselia adalah teknik perlombongan data di mana algoritma yang digunakan untuk membangunkan model adalah berasaskan kepada data yang mempunyai input dan output. Algoritma tersebut akan mempelajari fungsi untuk memetakan input dan output tersebut sebelum ia membuat ramalan ke atas data yang tidak pernah dilihat dan dipelajari. Set data siri masa khususnya bagi ramalan univariat tidak berada dalam format pembelajaran berselia dan perlu distrukturkan semula supaya menjadi data yang mempunyai ciri pembelajaran berselia. Kaedah yang biasa digunakan dalam analisis dan ramalan melibatkan data siri masa adalah kaedah tetingkap yang menyusun dan menstrukturkan semula susunan keseluruhan data supaya menjadi pembelajaran berselia. Melalui kaedah ini, satu tempoh masa harga saham yang lepas digunakan sebagai input untuk menentukan harga pada hari atau tempoh berikutnya. Berbanding menggunakan keseluruhan harga yang lepas, kaedah ini dapat mempercepatkan proses latihan algoritma dan memastikan corak serta harga terkini sahaja digunakan untuk melatih model ramalan dan memastikan ketepatan ramalan dapat dibuat (Hsu et al. 2016).

Kaedah tettingkap adalah pendekatan yang digunakan bagi menukar persembahan data siri masa kepada data dalam format pembelajaran berselia. Ini bagi membolehkan data tersebut digunakan untuk melatih algoritma pembelajaran mesin atau pembelajaran mendalam sebelum ia mampu membuat ramalan harga saham. Kaedah ini menggunakan data dari sela masa yang lepas sebagai fitur input untuk menentukan dan meramal harga saham pada sela masa yang berikutnya. Tettingkap ini akan digerakkan pada setiap sela bagi membentuk satu jujukan data baharu. Bilangan sela masa yang ditetapkan adalah merujuk kepada saiz tettingkap tersebut. Bagi tujuan kajian ini, sela masa selama 10 hari lepas digunakan sebagai saiz tettingkap seperti yang telah dilakukan oleh Liu & Long (2020), Nikou et al. (2019) dan Yong et al. (2017) dalam kajian mereka. Selain itu, sela masa selama 30 dan 60 hari juga turut digunakan untuk mempelbagaikan set data dan melihat perbezaan prestasi ramalan berasaskan saiz tettingkap yang ditetapkan.

Bagi saiz tettingkap sebanyak 10 hari, data dari sela masa pertama hingga sela masa ke-10 akan digunakan sebagai fitur input manakala data pada sela masa ke-11 akan menjadi nilai sasaran. Tettingkap ini akan digerakkan pada sela masa seterusnya di mana data dari sela masa ke-2 hingga ke-11 digunakan sebagai fitur input baharu dan data pada sela masa ke-12 akan menjadi nilai sasaran. Proses ini diteruskan sehingga sela masa yang terakhir di mana keseluruhan fitur-fitur input bersama nilai sasaran akan membentuk satu struktur data baharu dalam format pembelajaran berselia. Bagi saiz tettingkap sebanyak 30 dan 60 hari pula, pendekatan yang sama digunakan di mana tempoh tersebut digunakan bagi menentukan bilangan fitur input yang akan digunakan dan nilai sasaran bagi membentuk data dalam format pembelajaran berselia. Rajah 3.6 di bawah menunjukkan konsep tettingkap yang digunakan untuk mengolah dan menukar data siri masa bagi menjadikan set data tersebut sebagai pembelajaran berselia.



Rajah 3.6 Kaedah tetingkap bagi menukar data siri masa kepada set data dalam format pembelajaran berselia

Disebakan kajian ini memberi fokus kepada ramalan harga tutup saham, maka fitur Close merupakan satu-satunya fitur yang akan digunakan dalam membuat ramalan univariat. Fitur tersebut akan digunakan dalam kaedah tetingkap untuk membentuk data pembelajaran berselia bagi membuat ramalan univariat ke atas harga tutup saham. Bagi ramalan multivariat pula, kesemua fitur iaitu Open, Close, High, Low dan Volume akan digunakan dalam kaedah tetingkap tersebut untuk membentuk data pembelajaran berselia. Kaedah tetingkap ini juga akan menjadi kaedah utama dalam kajian ini bagi membentuk beberapa set data berasaskan kepada saiz tetingkap dan teknik ramalan sama ada menggunakan ramalan univariat atau multivariat. Penyediaan set data ini akan dihuraikan dengan lebih lanjut dalam sub tajuk yang berikutnya.

#### 3.4.4 Penyediaan Set Data

Setelah semua aktiviti pra-pemprosesan yang dinyatakan sebelum ini selesai, set data akan disediakan sebagai data akhir yang akan digunakan untuk membangunkan dan menguji model ramalan. Penyediaan set data ini dibuat berasaskan kepada variasi tetapan saiz tetingkap dan pendekatan ramalan univariat atau multivariat. Bagi ramalan harga saham pada hari berikutnya, sebanyak enam set data akan digunakan bagi setiap data saham yang melibatkan tetapan saiz tetingkap sebanyak 10, 30 dan 60 hari serta

kaedah ramalan menggunakan pendekatan univariat dan multivariat.. Bagi ramalan jangka panjang yang melibatkan ramalan pada hari ke-3, hari ke-5 dan hari ke-7 pula, sebanyak 6 set data juga akan digunakan bagi setiap data saham yang melibatkan tetapan saiz tettingkap sebanyak 10, 30 dan 60 hari. Kesemua set data tersebut disediakan mengikut tetapan saiz tettingkap yang berbeza serta pendekatan univariat dan multivariat.

Penyediaan set data yang pelbagai ini bertujuan untuk melihat perbezaan keupayaan dan prestasi model ramalan berdasarkan kepada saiz tettingkap dan pendekatan univariat serta multivariat yang dipilih bagi setiap set data tersebut. Jadual 3.4 di bawah menunjukkan kesemua set data yang akan digunakan di mana perbezaan setiap set adalah berasaskan kepada tempoh sela masa atau saiz tettingkap dan jenis pendekatan ramalan yang digunakan iaitu univariat dan multivariat.

Jadual 3.4 Set data serta variasi tetapan yang digunakan

Set Data	Fitur Input	Jenis Ramalan	Saiz Tettingkap
Set Data 1	Close	Univariat	10 hari
Set Data 2	Close	Univariat	30 hari
Set Data 3	Close	Univariat	60 hari
Set Data 4	Open, Close, High, Low, Volume	Multivariat	10 hari
Set Data 5	Open, Close, High, Low, Volume	Multivariat	30 hari
Set Data 6	Open, Close, High, Low, Volume	Multivariat	60 hari

Secara keseluruhannya, setiap data saham akan mempunyai 6 set data yang berbeza yang mana ini menjadikan jumlah set data bagi kedua-dua data saham adalah sebanyak 12 set data.

### 3.5 PEMODELAN

Fasa pemodelan merupakan peringkat di mana model ramalan akan dibangunkan menggunakan algoritma yang telah dipilih dan data-data yang telah diproses bagi memenuhi matlamat dan objektif asal kajian. Bagi projek atau kajian yang melibatkan pembelajaran mesin, fasa ini merupakan fasa utama yang akan menentukan keputusan keseluruhan yang mempengaruhi objektif yang telah ditetapkan. Sebanyak 6 algoritma berbeza akan digunakan bagi membangunkan model ramalan. Tiga model pembelajaran mesin tradisional iaitu DTR, RFR dan SVR akan dibangunkan sebagai model dasar bagi

tujuan perbandingan prestasi. Pemilihan tiga model ini dibuat berdasarkan kepada Mehtab et al. 2021, Nabipour et al. 2020 dan Nikou et al. 2019 yang menggunakan model-model ini sebagai perbandingan dalam kajian mereka. Bagi pembelajaran mendalam pula, MLP, CNN dan LSTM akan digunakan sebagai algoritma untuk membangunkan model ramalan.

### 3.5.1 Pecahan Data Latihan dan Pengujian

Bagi membangunkan model ramalan menggunakan teknik pembelajaran mesin dan pembelajaran mendalam, ia perlu dilatih terlebih dahulu untuk memahami dan mempelajari data saham sebelum mampu membuat ramalan. Data yang telah diproses sebelum ini akan diberikan kepada algoritma supaya ia boleh mencari dan mempelajari corak, hubungan dan fungsi untuk memetakan harga saham yang lepas sebagai input dengan nilai sasaran sebagai output. Namun model tersebut juga perlu diuji dengan data yang tidak pernah dilihat serta difahami olehnya supaya keupayaan dan prestasi sebenarnya dapat dibuktikan serta dinilai. Oleh itu, data kajian yang telah diproses perlu dipecahkan dan dibahagikan kepada dua bahagian utama iaitu data bagi tujuan melatih model dan data bagi tujuan menguji keupayaan model.

Untuk tujuan tersebut Nikou et al. 2019 adalah dirujuk di mana sebanyak 80% daripada keseluruhan setiap set data akan digunakan sebagai data latihan manakala 20% bakinya untuk tujuan pengujian. Peratus pecahan ini dipilih kerana model perlu dilatih dengan saiz data yang lebih besar berbanding dengan saiz data untuk pengujian supaya ia dapat mempelajarinya dengan lebih efisien dan seterusnya mampu membuat ramalan dengan lebih tepat. Jadual 3.5 di bawah menunjukkan pecahan data serta jumlah rekod yang akan digunakan untuk tujuan latihan dan pengujian model.

Jadual 3.5 Peratusan pecahan set data serta jumlah rekod yang akan digunakan bagi tujuan latihan dan pengujian

	Peratus	Jumlah Rekod
Latihan	80%	987
Pengujian	20%	246

Pecahan set data ini adalah proses awal untuk mengasingkan data bagi tujuan latihan dan juga pengujian. Bagi permasalahan data siri masa, transformasi ke atas data latihan



dan pengujian ini masih perlu dilakukan menggunakan kaedah tetingkap agar ianya dapat dipersembahkan dalam format pembelajaran berselia. Oleh itu, jumlah rekod bagi data latihan untuk setiap set data akan berbeza bergantung kepada saiz tetingkap. Ini disebabkan kaedah tersebut akan mengambil data pada sela masa yang awal mengikut saiz tetingkap yang ditetapkan untuk membentuk satu rekod baharu. Setelah transformasi data menggunakan kaedah tetingkap dibuat, jumlah rekod bagi data latihan akan berkurang mengikut saiz tetingkap. Bagi saiz tetingkap 10 hari, jumlah rekod latihan akan berkurang sebanyak 10 hari menjadikan jumlah baharu sebanyak 977. Saiz tetingkap 30 dan 60 hari pula akan membentuk data latihan dengan jumlah baharu sebanyak 957 dan 927.

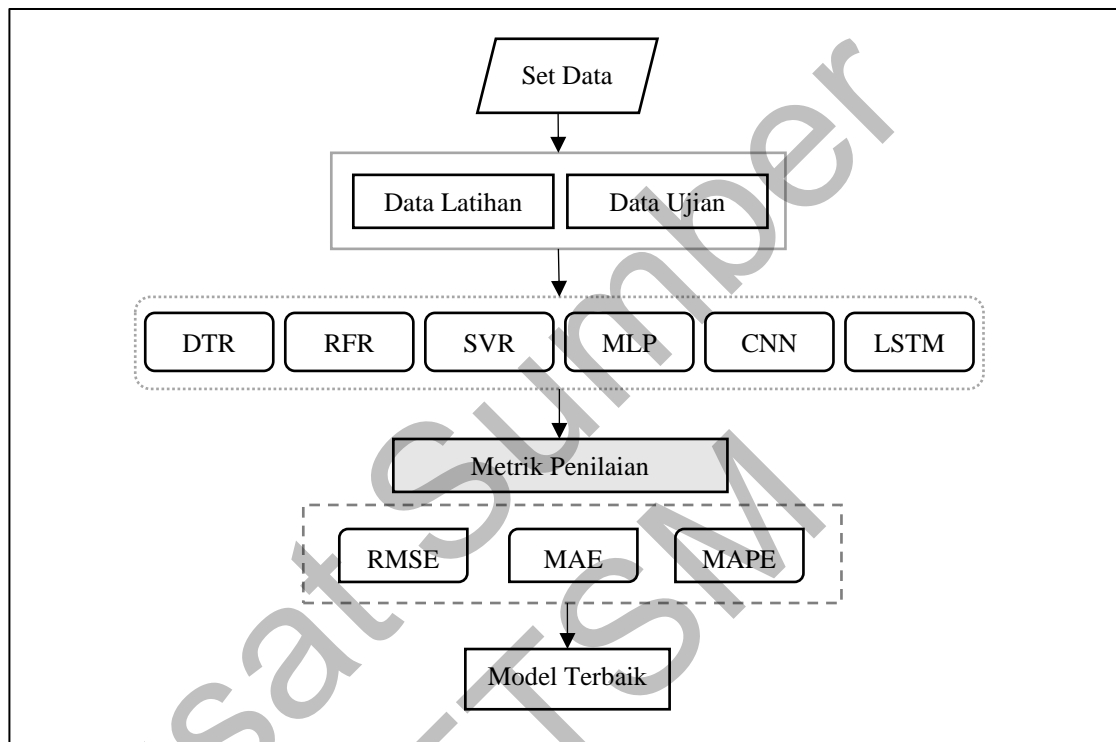
### 3.5.2 Pembangunan Model Ramalan

Setelah semua data siap diproses dan disediakan mengikut set-set data yang telah dinyatakan sebelum ini, pembangunan model ramalan akan dilakukan. Kesemua set data yang dinyatakan sebelum ini akan digunakan ke atas setiap algoritma yang telah dipilih bagi membangunkan model ramalan. Setiap model pula akan membuat ramalan harga saham pada hari ke-1, hari ke-3, hari ke-5 dan hari ke-7 menggunakan pendekatan univariat dan juga multivariat. Kombinasi set-set data bagi kedua-dua data saham serta sasaran dan tempoh ramalan ini akan melibatkan sebanyak 288 ramalan yang melibatkan tiga algoritma pembelajaran mesin tradisional dan tiga algoritma pembelajaran mendalam. Pecahan pembangunan model ramalan yang dibuat adalah seperti di dalam Jadual 3.6 di bawah.

Jadual 3.6 Bilangan model bagi setiap algoritma untuk kesemua tempoh ramalan

Algoritma	Axiata Group		Petronas Gas	
	Univariat	Multivariat	Univariat	Multivariat
DTR	12	12	12	12
RFR	12	12	12	12
SVR	12	12	12	12
MLP	12	12	12	12
CNN	12	12	12	12
LSTM	12	12	12	12
<b>Jumlah</b>	<b>72</b>	<b>72</b>	<b>72</b>	<b>72</b>

Secara ringkasnya, setiap set data yang telah ditetapkan akan digunakan ke atas setiap algoritma bagi membentuk model ramalan untuk setiap tempoh ramalan. Hasil output dan ramalan akan dinilai menggunakan metrik penilaian yang ditetapkan bagi menentukan prestasi dan keupayaan setiap model tersebut. Rajah 3.7 di bawah menunjukkan secara umum proses pembangunan model-model ramalan tersebut.



Rajah 3.7 Proses pembangunan model-model ramalan

Setiap model ramalan ini mempunyai struktur, rekabentuk dan kemampuan yang tersendiri. Pembangunan setiap model ini juga memerlukan tetapan atau konfigurasi yang berbeza mengikut kehendak atau keperluan penggunaannya.

**a. Model Regresi Pokok Keputusan (DTR)**

Algoritma ini membuat tugas regresi melalui struktur yang berasaskan kepada hierarki dan pecahan dahan pokok. Melalui struktur tersebut, set data akan dipecahkan kepada subset yang lebih kecil dan terperinci. Pecahan cabang struktur pokok tersebut ditentukan oleh algoritma ini melalui Min Ralat Kuasa Dua (MSE). Algoritma ini akan mempelajari struktur tersebut secara terperinci dan menggunakannya untuk meramal

nilai sasaran pada masa hadapan. Berdasarkan kepada fitur input daripada set data saham yang diberikan kepada algoritma ini, ia akan membuat pengiraan dan pecahan cabang berdasarkan kepada nilai MSE sehingga ia mendapat nilai ramalan yang disasarkan. Konfigurasi parameter yang digunakan bagi pembangunan model DTR untuk kajian ini adalah seperti dalam Jadual 3.7 di bawah.

Jadual 3.7 Konfigurasi parameter bagi pembangunan model DTR

Nama Parameter	Konfigurasi
random_state	0
splitter	best
criterion	squared_error
max_leaf_node	None
max_depth	None

**b. Model Regresi Hutan Rawak (RFR)**

Algoritma ini menggunakan teknik penggabungan di mana ia menggabungkan beberapa struktur pokok keputusan untuk mendapatkan nilai ramalan yang lebih tepat. Berbeza dengan regresi pokok keputusan yang menggunakan hanya satu struktur pokok untuk mendapatkan keputusan akhir, hutan rawak menggunakan gabungan bilangan pokok yang banyak dan mengambil purata semua keputusan yang diperolehi daripada setiap pokok untuk menentukan satu output akhir yang lebih tepat. Penilaian algoritma ini dibuat berdasarkan kepada pelbagai keputusan yang dihasilkan daripada setiap struktur pokok. Bagi tujuan pembangunan model ramalan, sebanyak 100 bilangan pokok ditetapkan melalui parameter `n_estimator` pada model tersebut. Purata keputusan daripada setiap pokok tersebut akan menjadi nilai ramalan harga saham yang akan dibuat oleh model tersebut. Konfigurasi parameter yang digunakan bagi pembangunan model RFR untuk kajian ini adalah seperti dalam Jadual 3.8 di bawah

Jadual 3.8 Konfigurasi parameter bagi pembangunan model RFR

Nama Parameter	Konfigurasi
random_state	0
n_estimator	100
criterion	squared_error
max_leaf_node	None
max_depth	None

### c. Model Regresi Sokongan Vektor (SVR)

SVR adalah algoritma yang menggunakan konsep dan asas yang sama digunakan oleh SVM dalam tugas pengelasan. Ia akan cuba mendapatkan satu garisan pemisah terbaik yang dikenali sebagai 'hyperplane' untuk mengasingkan data mengikut kluster yang bersesuaian. Hyperplane merupakan suatu sempadan yang akan memisahkan data kepada segmen yang berasingan untuk digunakan bagi meramal nilai output. Bagi permasalahan data yang tidak linear, kernel digunakan bagi membolehkan hyperplane yang lebih dinamik digunakan dalam dimensi yang lebih besar. Ia akan mengubah data yang tidak dapat dipisahkan secara linear dengan menambah dimensi kepada data tersebut supaya ia dapat disegmenkan menggunakan hyperplane. Dalam permasalahan ramalan harga saham ini, SVR akan memetakan fungsi melalui anggaran daripada input yang untuk menjana nilai output iaitu harga ramalan. Disebabkan data saham adalah data yang tidak linear, kernel RBF digunakan dalam pembangunan model ramalan ini. Ia membolehkan hyperlane yang lebih dinamik dibentuk dan memastikan nilai output atau ramalan yang lebih tepat dihasilkan. Konfigurasi parameter yang digunakan bagi pembangunan model SVR untuk kajian ini adalah seperti dalam Jadual 3.9 di bawah.

Jadual 3.9 Konfigurasi parameter bagi pembangunan model SVR

Nama Parameter	Konfigurasi
kernel	rbf
gamma	scale
shrinking	True
max_iter	-1

### d. Model Multilayer Perceptron (MLP)

Rekabentuk algoritma ini menggunakan rangkaian neural yang mempunyai satu lapisan input, satu lapisan output dan sekurang-kurangnya satu lapisan tersembunyi. Setiap lapisan tersembunyi akan mempunyai neuron yang memproses input yang diterima berasaskan kepada pemberat dan kecenderungan yang ditetapkan. Setiap neuron pula mempunyai hubungan terus dengan kesemua neuron daripada lapisan sebelumnya dan rangkaian neural yang bersambung secara penuh di antara kesemua neuron tersebut. Bagi pembangunan model ramalan ini, sebanyak dua lapisan tersembunyi digunakan

bersama satu lapisan input dan satu lapisan output. Ini akan membentuk struktur pembelajaran mendalam yang akan digunakan untuk meramal harga saham. Sebanyak 200 neuron pula ditetapkan pada setiap lapisan tersembunyi yang akan membaca dan memproses input yang diberikan sebelum dihantar kepada lapisan output. Konfigurasi parameter yang digunakan bagi pembangunan model MLP untuk kajian ini adalah seperti dalam Jadual 3.10 di bawah.

Jadual 3.10 Konfigurasi parameter bagi pembangunan model MLP

Nama Parameter	Konfigurasi
nodes	200
epochs	150
batch_size	24
activation	relu
optimizer	adam
loss	mse
learning_rate	0.001

e. **Model Rangkaian Neural Berlingkaran (CNN)**

Algoritma ini dibentuk berasaskan kepada tiga lapisan utama iaitu lapisan berlingkaran, lapisan pengumpulan dan lapisan berhubung sepenuhnya. Berbeza dengan rekabentuk MLP, CNN mempunyai lapisan berlingkaran dan lapisan pengumpulan yang berfungsi mengumpul maklumat penting daripada input yang diberikan sebelum dihantar kepada lapisan berhubung sepenuhnya untuk dikeluarkan sebagai output ramalan. Bagi pembangunan model ramalan menggunakan algoritma ini, sebanyak dua lapisan berlingkaran digunakan bagi membentuk struktur pembelajaran mendalam. Ia juga bagi membolehkan algoritma ini memilih dan menentukan maklumat serta intipati yang penting untuk digunakan dalam membuat ramalan. Saiz matriks yang ditetapkan untuk menapis dan memilih intipati maklumat adalah sebanyak tiga dan ditetapkan melalui parameter kernel di dalam algoritma ini. Konfigurasi parameter yang digunakan bagi pembangunan model CNN untuk kajian ini adalah seperti dalam Jadual 3.11 di bawah.

Jadual 3.11 Konfigurasi parameter bagi pembangunan model CNN

Nama Parameter	Konfigurasi
filters	256
kernel	3
epochs	100
batch_size	32
pool_size	2
activation	relu
optimizer	adam
loss	mse
learning_rate	0.001

#### f. Model Ingatan Jangka Pendek Panjang (LSTM)

Algoritma ini mempunyai struktur asas yang sama dengan rangkaian neural dengan satu lapisan input, lapisan tersembunyi dan lapisan output. Namun perbezaan utamanya adalah pada lapisan tersembunyi di mana ia mempunyai struktur blok memori atau sel yang boleh menguruskan ingatan sama ada perlu menyimpan atau membuang data-data lama mengikut keperluan. Ia memberikan kelebihan kepada algoritma ini dalam memahami dan membuat ramalan menggunakan data-data yang berjajukan dan siri masa seperti data saham yang mempunyai hubungan di antara satu dan lain. Bagi pembangunan model ramalan ini, sebanyak 2 lapisan tersembunyi ditetapkan pada algoritma ini bagi membentuk lapisan mendalam bersama dengan lapisan input dan juga output. Sebanyak 100 nod atau blok memori ditetapkan pada setiap lapisan tersembunyi bagi membolehkan algoritma ini mempelajari dan memahami dengan lebih efisien data saham yang digunakan sebelum ia mampu membuat ramalan. Hasil daripada proses di dalam lapisan tersembunyi tersebut akan dihantar kepada lapisan output untuk dikeluarkan sebagai harga ramalan yang dikehendaki. Konfigurasi parameter yang digunakan bagi pembangunan model LSTM untuk kajian ini adalah seperti dalam Jadual 3.12 di bawah.

Jadual 3.12 Konfigurasi parameter bagi pembangunan model LSTM

Nama Parameter	Konfigurasi
nodes	100
epochs	200
	bersambung...

...sambungan	
batch_size	32
activation	relu
optimizer	adam
loss	mse
learning_rate	0.001

---

### 3.6 PENILAIAN

Setelah model ramalan dilatih dan diuji, penilaian ke atas keupayaan dan prestasi ramalannya akan dibuat serta dibandingkan bagi membolehkan analisis terperinci dibuat. Ini bagi menentukan jenis algoritma serta set data yang mampu memberikan prestasi paling tinggi dalam membuat kedua-dua jenis ramalan. Selain itu, keupayaan algoritma pembelajaran mendalam untuk meramal pasaran saham juga dapat diukur serta dinilai dengan lebih tepat selaras dengan objektif asal kajian yang telah ditetapkan. Oleh itu, beberapa metrik penilaian akan digunakan untuk mengukur prestasi setiap model ramalan yang telah dibangunkan.

Punca Min Ralat Kuasa Dua (RMSE), Min Ralat Mutlak (MAE) dan Min Peratus Ralat Mutlak (MAPE) adalah tiga metrik penilaian yang digunakan untuk mengukur keupayaan dan ketepatan model ramalan meramal harga saham. Kesemua metrik penilaian ini dipilih sebagai pengukur prestasi berdasarkan kepada penggunaannya yang meluas dan menjadi pilihan dalam pelbagai kajian ramalan pasaran saham (Jiang 2021; Bustos & Pomares-Quimbaya 2020; Gandhmal & Kumar 2019). Selain itu, ketiga-tiga metrik penilaian ini adalah pilihan yang sering digunakan dalam tugas regresi bagi mengukur ketepatan nilai ramalan berbanding dengan nilai sebenar. Secara asasnya, ketiga-tiga metrik penilaian ini menggunakan ralat ramalan atau perbezaan nilai sebenar dengan nilai ramalan sebagai ukuran ketepatan. Oleh itu nilai metrik penilaian yang rendah menunjukkan prestasi ramalan yang baik berbanding dengan nilai metrik penilaian yang lebih tinggi.

#### 3.6.1 Punca Min Ralat Kuasa Dua (RMSE)

RMSE merupakan metrik penilaian yang menjadikan sisihan piawai pada ralat ramalan di dalam tugas regresi sebagai ukuran. Ralat ramalan ini menunjukkan jurang atau

jarak di antara nilai sebenar dengan nilai yang diramal oleh model serta bagaimana ia tersebar di sekelilingnya. Secara asasnya, RMSE mengira nilai kuasa dua bagi purata perbezaan di antara nilai sebenar dan nilai ramalan sebelum mencari semula punca kuasa dua bagi nilai tersebut. Oleh itu, nilai RMSE yang semakin mengecil menunjukkan perbezaan nilai ramalan dan sebenar yang tidak terlalu besar dan menghampiri nilai sebenarnya. Kiraan RMSE adalah seperti berikut:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Ramalan_i - Sebenar_i)^2}{N}}$$

di mana  $Ramalan_i$  merujuk kepada nilai ramalan pada data  $i$ ,  $Sebenar_i$  merujuk kepada nilai sebenar pada data  $i$  dan  $N$  merujuk kepada jumlah set data pengujian.

### 3.6.2 Min Ralat Mutlak (MAE)

MAE adalah metrik penilaian yang menjadikan perbezaan di antara nilai ramalan dan nilai sebenar sebagai ukuran ketepatan. Perbezaan tersebut akan dijadikan sebagai nilai mutlak bagi mengelakkan nilai negatif diambil kira sebagai pengiraan. MAE akan mengambil purata setiap perbezaan di dalam set ramalan dan dijadikan sebagai nilai ukuran. Nilai MAE yang kecil menunjukkan model menghasilkan ramalan yang mempunyai ralat yang sedikit berbanding dengan nilai asal. Pengiraan MAE dapat diringkaskan melalui formula seperti berikut:

$$MAE = \frac{\sum_{i=1}^N |Ramalan_i - Sebenar_i|}{N}$$

di mana  $Ramalan_i$  merujuk kepada nilai ramalan pada data  $i$ ,  $Sebenar_i$  merujuk kepada nilai sebenar pada data  $i$  dan  $N$  merujuk kepada jumlah set data pengujian.

### 3.6.3 Min Peratus Ralat Mutlak (MAPE)

MAPE pula adalah metrik penilaian yang menggunakan perbezaan di antara nilai sebenar dan ramalan dalam bentuk peratus sebagai ukuran ketepatan ramalan. Kelebihan MAPE berbanding RMSE dan MAE adalah ianya dipersembahkan dalam bentuk peratusan di mana nilai tersebut lebih mudah difahami. Ia juga sering digunakan sebagai metrik yang menggambarkan kadar ketepatan ramalan disebabkan ianya



dinyatakan dalam bentuk peratusan. Sebagai contoh, model yang mendapat MAPE dengan nilai 2% boleh digambarkan sebagai model yang mempunyai ketepatan sebanyak 98%. Nilai MAPE yang semakin kecil menunjukkan ralat yang sedikit di antara nilai ramalan berbanding nilai sebenar. Pengiraan MAPE adalah seperti formula berikut:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{Ramalan_i - Sebenar_i}{Sebenar_i} \right| * 100$$

di mana  $Ramalan_i$  merujuk kepada nilai ramalan pada data  $i$ ,  $Sebenar_i$  merujuk kepada nilai sebenar pada data  $i$  dan  $N$  merujuk kepada jumlah set data pengujian.

### 3.7 KESIMPULAN

Pelaksanaan kajian ini secara umumnya menjadikan pendekatan CRISP-DM sebagai asas dan kerangka utama metodologi kajian. Ia bersesuaian dengan skop kajian ini yang melibatkan aktiviti perlombongan data khususnya dalam meramal harga saham daripada data siri masa saham yang lepas. Perincian kepada aktiviti di dalam fasa model CRISP-DM ini pula dilaksanakan berpandukan kepada rujukan daripada kajian kepustakaan yang lepas sebagai asas dan panduan untuk melaksanakan kajian ini. Jiang 2021 dan Bustos & Pomares-Quimbaya 2020 dijadikan asas dalam memilih beberapa pendekatan dan tetapan ke atas metodologi kajian ini. Selain itu, kajian-kajian seperti yang dilaksanakan oleh Nikou et al. 2019, Mehtab et al. 2020 dan Nabipour et al. 2020 digunakan sebagai rujukan utama supaya keputusan uji kaji nanti dapat dibuat dan dianalisis secara lebih mudah dan dibuat perbandingan.

## **BAB IV**

### **KEPUTUSAN DAN PENILAIAN**

#### **4.1 PENGENALAN**

Bab ini akan menerangkan pemodelan yang dicadangkan bagi membuat ramalan harga saham, analisis serta penilaian ke atas keputusan uji kaji yang dihasilkan oleh kesemua model ramalan. Seperti yang telah dinyatakan di dalam bab sebelum ini, sebanyak tiga model asas yang menggunakan teknik pembelajaran mesin tradisional iaitu Regresi Pokok Keputusan (DTR), Regresi Hutan Rawak (RFR) dan Regresi Sokongan Vektor (SVR) dibangunkan sebagai perbandingan ke atas tiga algoritma pembelajaran mendalam iaitu MLP, CNN dan LSTM dalam membuat ramalan harga saham. Penilaian prestasi dan keupayaan kesemua model ini dibuat berdasarkan kepada tiga metrik penilaian iaitu RMSE, MAE dan MAPE bagi melihat dan menentukan ketepatan ramalan yang dibuat. Penilaian tersebut juga akan dibuat bagi melihat keupayaan dan kemampuan teknik pembelajaran mendalam berbanding pembelajaran mesin tradisional dalam meramal harga saham.

#### **4.2 RAMALAN UNIVARIAT**

Pendekatan univariat dalam membuat ramalan harga saham adalah kaedah yang berasaskan kepada pemerhatian satu fitur sahaja dan tidak melibatkan fitur-fitur lain sebagai input kepada model. Fitur tersebut adalah fitur yang sama dengan sasaran ramalan iaitu harga tutup saham. Pendekatan ini digunakan ke atas kedua-dua data saham dan juga pelbagai tetapan saiz tettingkap bagi meramal harga tutup saham pada hari pertama, ketiga, kelima dan ketujuh. Hasil dan keputusan daripada uji kaji ini dipecahkan mengikut tetapan set-set data yang berbeza supaya analisis dan penilaian prestasi dapat dibuat ke atas setiap keputusan tersebut.

#### 4.2.1 Ramalan Harga Saham Satu Hari Ke Depan

##### a. Saiz Tetingkap 10 Hari

Hasil penilaian daripada uji kaji yang dijalankan ke atas kesemua model dengan menggunakan saiz tetingkap sebanyak 10 hari adalah seperti di dalam Jadual 4.1 di bawah.

Jadual 4.1 Analisis prestasi ramalan univariat dengan saiz tetingkap 10 hari

Algoritma	Axiata Group Berhad			Petronas Gas Berhad		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
DTR	0.1084	0.0829	2.2096	0.2394	0.1794	1.1198
RFR	0.0803	0.0595	1.5886	0.1806	0.1268	0.7873
SVR	0.0978	0.0769	2.0995	0.1849	0.1318	0.8209
MLP	0.0739	0.0555	1.4873	0.1677	0.1211	0.7543
CNN	0.0734	0.0557	1.4999	0.1646	0.1207	0.7521
LSTM	0.0692	0.0517	1.3861	0.1623	0.1171	0.7281

Uji kaji ini menunjukkan prestasi ramalan ketiga-tiga model pembelajaran mendalam iaitu MLP, CNN dan LSTM mengatasi keupayaan kesemua model pembelajaran mesin iaitu DTR, RFR dan SVR. LSTM merupakan model yang mempunyai prestasi ramalan yang paling baik bagi kedua-dua jenis data saham dengan nilai RMSE, MAE dan MAPE paling rendah berbanding kesemua model yang lain. Bagi saham Axiata Group, nilai RMSE, MAE dan MAPE yang diperolehi adalah serendah 0.0692, 0.0517 dan 1.3861 manakala bagi saham Petronas Gas pula adalah 0.1623, 0.1171 dan 0.7281. Model DTR pula merupakan model yang mendapat prestasi paling rendah dengan nilai RMSE, MAE dan MAPE pada kadar 0.1084, 0.0829 dan 2.2096 bagi saham Axiata Group manakala 0.2394, 0.1794 dan 1.1198 bagi saham Petronas Gas. Model RFR merupakan model yang mendapat prestasi terbaik di antara model pembelajaran mesin yang lain, namun ia masih gagal mengatasi prestasi ketiga-tiga model pembelajaran mendalam. Analisis ini juga menunjukkan bahawa ralat ramalan bagi saham Petronas Gas adalah lebih tinggi berbanding saham Axiata Group di mana nilai ketiga-tiga metrik penilaian bagi saham tersebut meningkat bagi kesemua model.

**b. Saiz Tetingkap 30 Hari**

Hasil penilaian daripada uji kaji yang dijalankan ke atas kesemua model dengan menggunakan saiz tetingkap sebanyak 30 hari adalah seperti di dalam Jadual 4.2 di bawah.

Jadual 4.2 Analisis prestasi ramalan univariat dengan saiz tetingkap 30 hari

Algoritma	Axiata Group Berhad			Petronas Gas Berhad		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
DTR	0.1288	0.1021	2.7436	0.2560	0.1971	1.2314
RFR	0.0785	0.0591	1.5814	0.1799	0.1302	0.8076
SVR	0.1044	0.0851	2.3155	0.2327	0.1874	1.1619
MLP	0.0750	0.0577	1.5529	0.2037	0.1686	1.0614
CNN	0.0771	0.0600	1.6199	0.1686	0.1245	0.7756
LSTM	0.0712	0.0538	1.4440	0.1599	0.1133	0.7096

Secara keseluruhannya, uji kaji ini menunjukkan ketiga-tiga model pembelajaran mendalam mendapat prestasi ramalan lebih baik daripada ketiga-tiga model pembelajaran mesin. Namun bagi saham Petronas Gas, model RFR mengatasi keupayaan model MLP dengan nilai RMSE, MAE dan MAPE pada kadar 0.1799, 0.1302 dan 0.8076. Uji kaji ini juga menunjukkan model pembelajaran mendalam LSTM mendapat prestasi ramalan yang paling baik bagi kedua-dua jenis data saham dengan nilai RMSE, MAE dan MAPE paling rendah berbanding kesemua model yang lain. Ia mendapat nilai RMSE, MAE dan MAPE sebanyak 0.0712, 0.0538 dan 1.4440 bagi saham Axiata Group manakala 0.1599, 0.1133 dan 0.7096 bagi saham Petronas Gas. Model DTR masih menjadi model dengan prestasi paling rendah berbanding kesemua model lain dengan nilai RMSE, MAE dan MAPE pada kadar 0.1288, 0.1021 dan 2.7436 bagi saham Axiata Group manakala 0.2560, 0.1971 dan 1.2314 bagi saham Petronas Gas. Model RFR pula adalah model pembelajaran mesin dengan prestasi paling baik berbanding lain. Analisis ini juga menunjukkan bahawa ralat ramalan bagi saham Petronas Gas Berhad adalah lebih tinggi berbanding saham Axiata Group. Selain itu, nilai RMSE, MAE dan MAPE bagi kesemua model juga meningkat berbanding dengan model yang menggunakan saiz tetingkap sebanyak 10 hari.

### c. Saiz Tetingkap 60 Hari

Hasil penilaian daripada uji kaji yang dijalankan ke atas kesemua model dengan menggunakan saiz tetingkap sebanyak 60 hari adalah seperti di dalam Jadual 4.3 di bawah.

Jadual 4.3 Analisis prestasi ramalan univariat dengan saiz tetingkap 60 hari

Algoritma	Axiata Group Berhad			Petronas Gas Berhad		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
DTR	0.1292	0.1002	2.6883	0.3387	0.2490	1.5554
RFR	0.0771	0.0580	1.5542	0.1923	0.1370	0.8507
SVR	0.1242	0.0994	2.6875	0.4252	0.3135	1.9298
MLP	0.0822	0.0643	1.7299	0.2372	0.1960	1.2234
CNN	0.0885	0.0697	1.8640	0.1671	0.1254	0.7844
LSTM	0.0729	0.0548	1.4747	0.1530	0.1111	0.6972

Uji kaji ini juga menunjukkan model pembelajaran mendalam LSTM mendapat prestasi ramalan yang paling baik bagi kedua-dua jenis data saham dengan nilai RMSE, MAE dan MAPE paling rendah berbanding kesemua model yang lain. Ia mendapat nilai RMSE, MAE dan MAPE sebanyak 0.0729, 0.0548 dan 1.4747 bagi saham Axiata Group manakala 0.1530, 0.1111 dan 0.6972 bagi saham Petronas Gas. Bagi set data ini, model DTR masih kekal dengan prestasi paling rendah berbanding kesemua model yang lain. Model RFR juga masih menjadi model pembelajaran mesin yang mendapat prestasi paling baik dan mengatasi model MLP bagi kedua-dua saham. Ia juga mempunyai nilai MAE yang lebih baik berbanding model CNN bagi walaupun mempunyai nilai RMSE yang agak tinggi bagi saham Petronas Gas. Peningkatan ralat ramalan bagi kesemua model juga agak ketara berbanding dua set data sebelum ini di mana nilai RMSE, MAE dan MAPE yang diperolehi meningkat naik. Analisis juga masih menunjukkan kesukaran kesemua model untuk mengurangkan ralat ramalan bagi saham Petronas Gas berbanding saham Axiata Group.

### d. Analisis Prestasi Model Ramalan

Prestasi ramalan secara keseluruhan oleh semua model dilihat agak baik dengan ralat ramalan yang kecil diperolehi bagi kedua-dua saham. Namun ketiga-tiga model

pembelajaran mendalam jelas menunjukkan prestasi ramalan yang lebih baik dibandingkan dengan prestasi ramalan model pembelajaran mesin. Ia dibuktikan dengan nilai RMSE, MAE dan MAPE yang diperolehi oleh kesemua model pembelajaran mendalam lebih kecil berbanding model pembelajaran mesin khususnya bagi ramalan harga saham Axiata Group. Bagi saham Petronas Gas pula, kesemua model menunjukkan peningkatan ralat ramalan jika dibandingkan dengan ramalan saham Axiata Group. Ia mungkin dipengaruhi oleh faktor pergerakan harga saham yang lebih rumit dan besar berbanding pergerakan harga saham Axiata Group.

Tetapan saiz tettingkap pula dilihat memberikan prestasi ramalan yang agak berbeza mengikut jenis data saham. Model LSTM mendapat prestasi ramalan terbaik bagi setiap saiz tettingkap untuk kedua-dua saham. Saiz tettingkap sebanyak 10 hari memberikan prestasi ramalan yang lebih baik bagi saham Axiata Group di mana peningkatan saiz kepada 30 dan 60 hari menurunkan sedikit prestasinya. Untuk saham Petronas Gas pula, saiz tettingkap sebanyak 60 hari memberikan prestasi ramalan yang lebih baik manakala penurunan saiz kepada 30 dan 10 hari dilihat menurunkan sedikit prestasinya. Rajah 4.1 dan Rajah 4.2 di bawah menunjukkan perbandingan nilai RMSE, MAE dan MAPE bagi model terbaik untuk setiap saiz tettingkap ke atas kedua-dua saham.

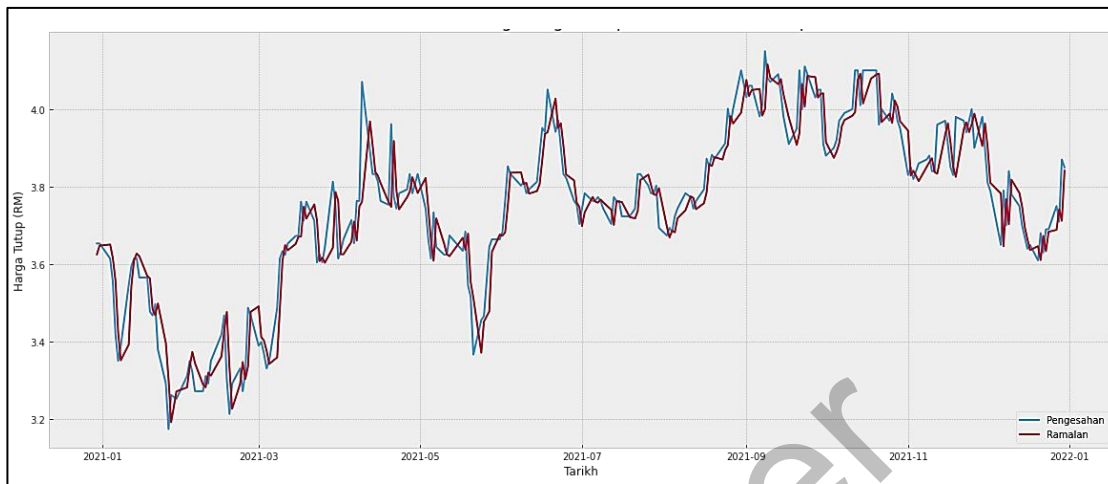


Rajah 4.1 Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Axiata Group

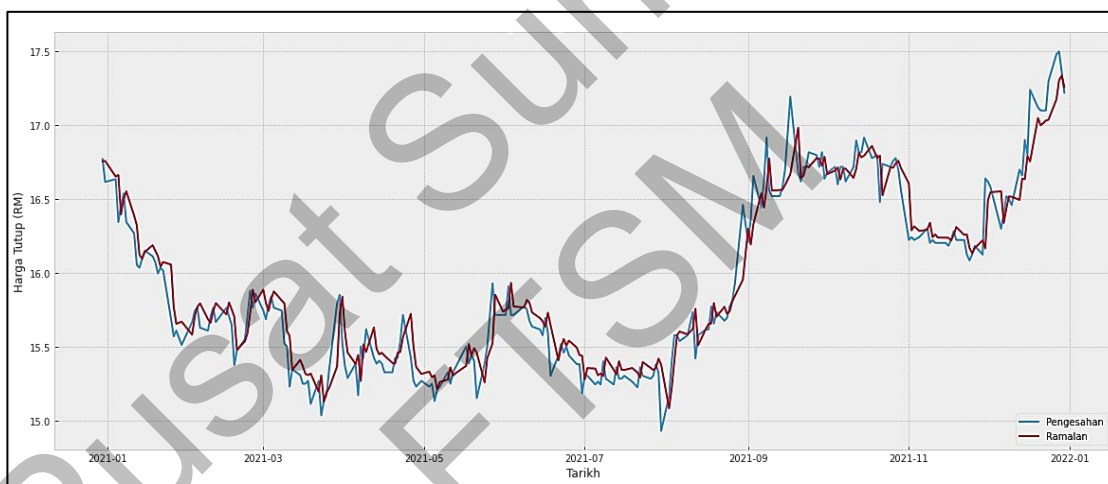


Rajah 4.2 Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Petronas Gas

Hasil daripada semua analisis yang dibuat, didapati model LSTM dengan tetapan saiz tettingkap sebanyak 10 hari adalah model terbaik bagi meramal harga saham Axiata Group untuk tempoh satu hari ke depan. Bagi saham Petronas Gas pula, model LSTM dengan tetapan saiz tettingkap sebanyak 60 hari merupakan model ramalan terbaik untuk tempoh satu hari ke depan. Perbandingan ramalan harga saham satu hari ke depan dengan harga tutup saham yang sebenar untuk kedua-dua data saham adalah seperti Rajah 4.3 dan Rajah 4.4 di bawah. Hasil ramalan ke atas kedua-dua harga tutup saham tersebut tidak menunjukkan perbezaan yang agak ketara dengan harganya yang sebenar dan masih mengikut corak pergerakan harga sebenar tersebut secara konsisten.



Rajah 4.3 Ramalan model LSTM dengan pendekatan univariat ke atas harga tutup saham Axiata Group untuk tempoh satu hari ke depan berbanding harga sebenar



Rajah 4.4 Ramalan model LSTM dengan pendekatan univariat ke atas harga tutup saham Petronas Gas untuk tempoh satu hari ke depan berbanding harga sebenar

## 4.2.2 Ramalan Harga Saham Tiga Hari Ke Depan

### a. Saiz Tetingkap 10 Hari

Hasil penilaian daripada uji kaji yang dijalankan ke atas kesemua model dengan menggunakan saiz tetingkap sebanyak 10 hari adalah seperti di dalam Jadual 4.4 di bawah.



Jadual 4.4 Analisis prestasi ramalan harga saham 3 hari ke depan dengan saiz tettingkap 10 hari

Algoritma	Axiata Group Berhad			Petronas Gas Berhad		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
DTR	0.1785	0.1423	3.8050	0.3509	0.2543	1.5818
RFR	0.1276	0.1020	2.7344	0.2479	0.1823	1.1320
SVR	0.1902	0.1501	4.1241	0.2871	0.2208	1.3828
MLP	0.1122	0.0896	2.4180	0.2348	0.1736	1.0798
CNN	0.1124	0.0903	2.4364	0.2401	0.1769	1.1013
LSTM	0.1100	0.0884	2.3734	0.2314	0.1685	1.0502

Secara keseluruhannya, uji kaji ini menunjukkan kesemua model pembelajaran mendalam iaitu MLP, CNN dan LSTM mendapat prestasi yang lebih baik berbanding dengan model pembelajaran mesin iaitu DTR, RFR dan SVR. Model LSTM merupakan model yang menunjukkan prestasi ramalan paling baik dengan nilai RMSE, MAE dan MAPE pada kadar 0.1100, 0.0884 dan 2.3734 bagi saham Axiata Group dan 0.2314, 0.1685 dan 1.0502 bagi saham Petronas Gas. Model DTR merupakan model yang mendapat prestasi ramalan paling rendah berbanding kesemua model lain dengan nilai RMSE, MAE dan MAPE pada kadar 0.1785, 0.1423 dan 3.8050 bagi saham Axiata dan 0.3509, 0.2543 dan 1.5818 bagi saham Petronas Gas. Bagi model pembelajaran mesin, RFR merupakan model dengan prestasi ramalan paling baik berbanding DTR dan SVR namun masih gagal mengatasi prestasi ramalan model pembelajaran mendalam. Analisis ini juga menunjukkan bahawa ralat ramalan bagi saham Petronas Gas adalah lebih tinggi berbanding saham Axiata Group di mana nilai ketiga-tiga metrik penilaian bagi saham tersebut meningkat bagi kesemua model.

#### b. Saiz Tetingkap 30 Hari

Hasil penilaian daripada uji kaji yang dijalankan ke atas kesemua model dengan menggunakan saiz tettingkap sebanyak 30 hari adalah seperti di dalam Jadual 4.5 di bawah.

Jadual 4.5 Analisis prestasi ramalan harga saham 3 hari ke depan dengan saiz tettingkap 30 hari

Algoritma	Axiata Group Berhad			Petronas Gas Berhad		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
DTR	0.2322	0.1808	4.8251	0.3538	0.2772	1.7322
RFR	0.1273	0.1033	2.7842	0.2467	0.1782	1.1053
SVR	0.1578	0.1294	3.5181	0.3413	0.2798	1.7375
MLP	0.1290	0.1037	2.8129	0.2670	0.2104	1.3094
CNN	0.1188	0.0944	2.5439	0.3019	0.2417	1.4986
LSTM	0.1120	0.0883	2.3844	0.2481	0.1906	1.1865

Uji kaji ini menunjukkan model LSTM dan CNN masih menjadi model dengan prestasi ramalan paling baik bagi saham Axiata Group. Model LSTM merupakan model terbaik bagi saham Axiata dengan nilai RMSE, MAE dan MAPE pada kadar 0.1120, 0.0883 dan 2.3844. Namun begitu, bagi saham Petronas Gas, ketiga-tiga model MLP, CNN dan LSTM menghadapi kesukaran untuk mengekalkan prestasi terbaik di mana model RFR mengatasi kesemua model lain dengan nilai RMSE, MAE dan MAPE pada kadar 0.2467, 0.1782 dan 1.1053. Model DTR merupakan model yang mendapat prestasi ramalan paling rendah berbanding kesemua model lain dengan nilai RMSE, MAE dan MAPE pada kadar 0.2322, 0.1808 dan 4.8251 bagi saham Axiata Group dan 0.3538, 0.2772 dan 1.7322 bagi saham Petronas Gas. Peningkatan ralat ramalan bagi kesemua model dapat dilihat dengan nilai RMSE, MAE dan MAPE yang lebih tinggi berbanding dengan uji kaji menggunakan saiz tettingkap sebanyak 10 hari. Namun hanya model RFR dilihat menunjukkan peningkatan prestasi berbanding uji kaji sebelum ini dengan sedikit penurunan nilai RMSE, MAE dan MAPE bagi kedua-dua saham.

### c. Saiz Tettingkap 60 Hari

Hasil penilaian daripada uji kaji yang dijalankan ke atas kesemua model dengan menggunakan saiz tettingkap sebanyak 60 hari adalah seperti di dalam Jadual 4.6 di bawah:

Jadual 4.6 Analisis prestasi ramalan harga saham 3 hari ke depan dengan saiz tettingkap 60 hari

Algoritma	Axiata Group Berhad			Petronas Gas Berhad		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
DTR	0.2628	0.1972	5.2976	0.4054	0.3119	1.9518
RFR	0.1204	0.0953	2.5779	0.2508	0.1884	1.1714
SVR	0.1697	0.1372	3.6961	0.5373	0.4136	2.5553
MLP	0.1479	0.1120	2.9992	0.3146	0.2469	1.5236
CNN	0.1220	0.0969	2.6093	0.2474	0.1835	1.1463
LSTM	0.1200	0.0941	2.5332	0.2422	0.1803	1.1208

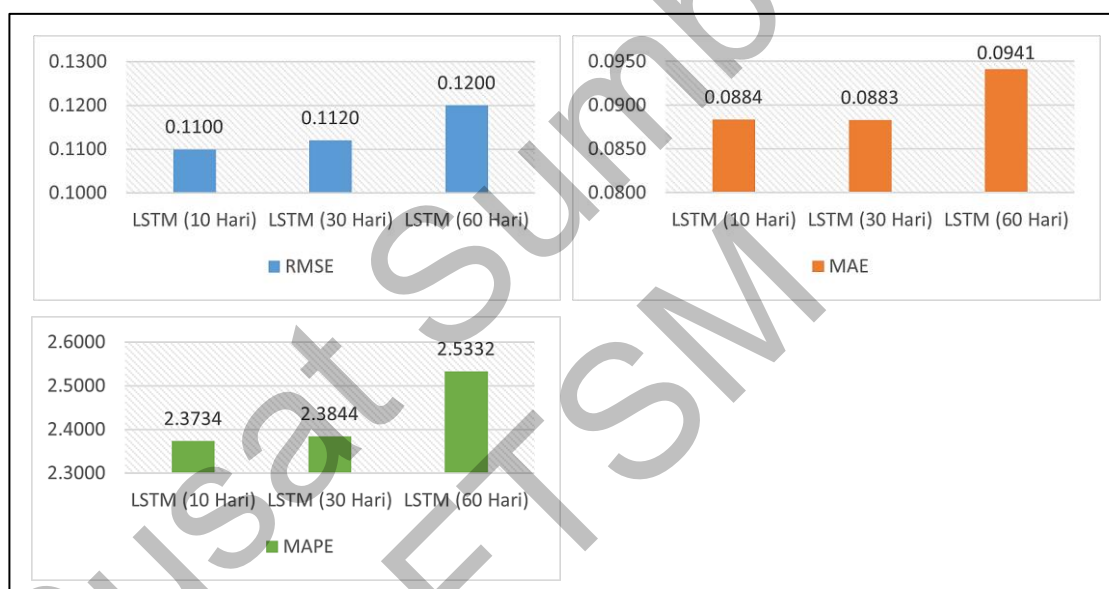
Keputusan uji kaji ini pula menunjukkan model LSTM mendapat prestasi ramalan paling baik bagi saham Axiata Group dan Petronas Gas. Ia mendapat nilai RMSE, MAE dan MAPE pada kadar 0.1200, 0.0941 dan 2.5332 bagi saham Axiata dan 0.2422, 0.1803 dan 1.1208 bagi saham Petronas Gas. LSTM masih mengekalkan prestasi ramalan paling baik di antara model pembelajaran mendalam yang lain. Model DTR pula menjadi model yang mendapat prestasi ramalan paling rendah berbanding kesemua model lain dengan nilai RMSE, MAE dan MAPE pada kadar 0.2628, 0.1972 dan 5.2976 bagi saham Axiata Group dan 0.4054, 0.3119 dan 1.9518 bagi saham Petronas Gas.

#### d. Analisis Prestasi Model Ramalan

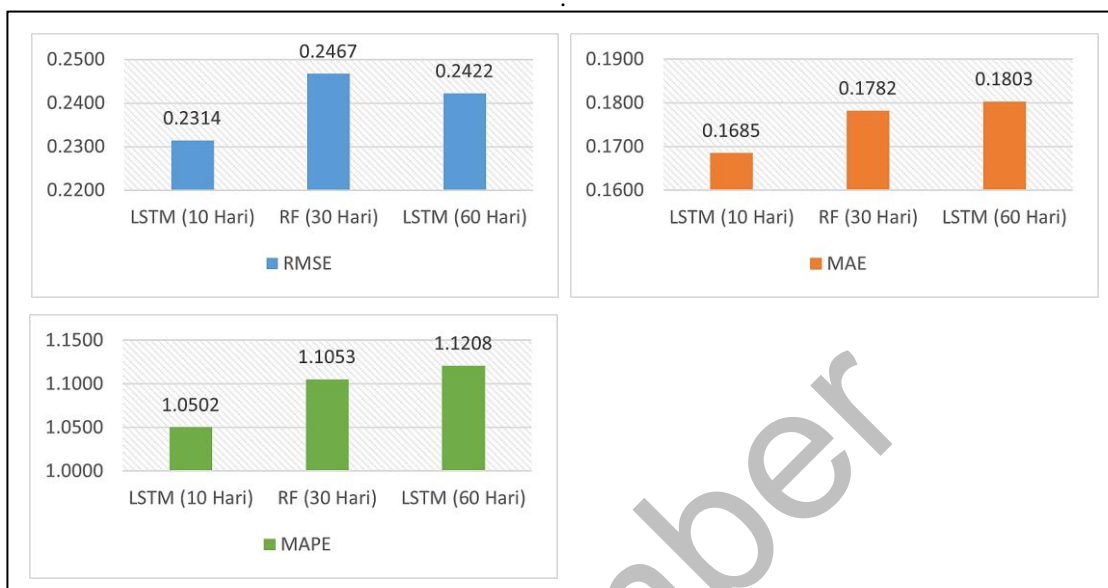
Secara keseluruhannya, model pembelajaran mendalam LSTM, CNN dan MLP merupakan model terbaik bagi ramalan harga untuk kedua-dua saham. Ini berdasarkan kepada metrik penilaian keseluruhan RMSE, MAE dan MAPE yang diperolehi oleh kesemua model tersebut yang lebih kecil berbanding model yang lain. Walaupun model RFR mendapat prestasi ramalan lebih baik bagi saham Petronas Gas pada saiz tettingkap 30 hari, namun prestasi ramalan terbaik secara keseluruhannya masih diperolehi oleh ketiga-tiga model pembelajaran mendalam khususnya model LSTM.

Pemilihan saiz tettingkap bagi kedua-dua saham dilihat mempengaruhi prestasi ramalan oleh setiap model. Bagi saham Axiata Group, model LSTM merupakan model yang mendapat prestasi paling baik bagi kesemua saiz tettingkap. Bagi saham Petronas Gas pula, prestasi terbaik bagi saiz tettingkap 10 dan 60 hari diperolehi oleh model

LSTM manakala RFR untuk saiz tettingkap 30 hari. Namun saiz tettingkap sebanyak 10 hari merupakan tetapan saiz yang paling efisien dan mampu memberikan prestasi ramalan paling baik bagi kesemua model khususnya model pembelajaran mendalam. Ini berdasarkan nilai RMSE, MAE dan MAPE yang rendah diperolehi pada saiz tersebut. Rajah 4.5 dan Rajah 4.6 menunjukkan perbezaan nilai RMSE, MAE dan MAPE bagi model-model yang mendapat prestasi paling baik untuk setiap saiz tettingkap bagi kedua-dua saham. Ia jelas menunjukkan saiz tettingkap sebanyak 10 hari bagi model LSTM merupakan tetapan yang mampu memberikan prestasi ramalan yang paling baik untuk kedua-dua saham berbanding saiz tettingkap yang lain.



Rajah 4.5 Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Axiata Group

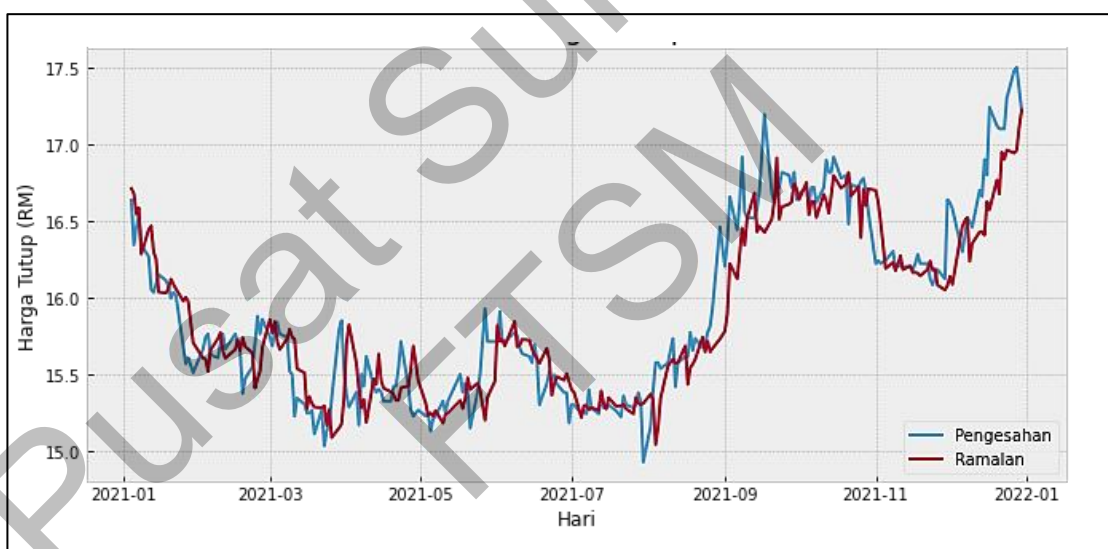


Rajah 4.6 Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Petronas Gas

Berdasarkan kepada semua analisis yang dibuat, model LSTM dengan saiz tettingkap sebanyak 10 hari adalah model terbaik bagi meramal harga saham Axiata Group dan Petronas Gas untuk tiga hari ke depan. Ini berdasarkan kepada nilai RMSE, MAE dan MAPE yang diperolehi oleh model ini adalah paling kecil berbanding dengan tetapan saiz tettingkap yang berlainan serta model-model lain. Perbandingan ramalan harga saham tiga hari ke depan dengan harga tutup saham yang sebenar untuk kedua-dua data saham yang dibuat oleh model ini adalah seperti Rajah 4.7 dan Rajah 4.8 di bawah. Hasil ramalan ke atas kedua-dua harga tutup saham tersebut dilihat mula menunjukkan jurang perbezaan yang agak ketara di antara harga ramalan dan juga harga yang sebenar. Namun ia masih berada dalam julat yang minimum dan masih mengikut corak pergerakan harganya secara konsisten.



Rajah 4.7 Ramalan model LSTM dengan pendekatan univariat ke atas harga tutup saham Axiata Group untuk tempoh tiga hari ke depan berbanding harga sebenar



Rajah 4.8 Ramalan model LSTM dengan pendekatan univariat ke atas harga tutup saham Petronas Gas untuk tempoh tiga hari ke depan berbanding harga sebenar

#### 4.2.3 Ramalan Harga Saham Lima Hari Ke Depan

##### a. Saiz Tetingkap 10 Hari

Hasil penilaian daripada uji kaji yang dijalankan ke atas kesemua model dengan menggunakan saiz tetingkap sebanyak 10 hari adalah seperti di dalam Jadual 4.7 di bawah:

Jadual 4.7 Analisis prestasi ramalan harga saham 5 hari ke depan dengan saiz tettingkap 10 hari

Algoritma	Axiata Group Berhad			Petronas Gas Berhad		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
DTR	0.2225	0.1727	4.6478	0.4421	0.3396	2.1128
RFR	0.1537	0.1248	3.3466	0.3005	0.2238	1.3909
SVR	0.2814	0.2279	6.2718	0.3524	0.2830	1.7781
MLP	0.1381	0.1095	2.9761	0.2514	0.1912	1.1901
CNN	0.1509	0.1261	3.3546	0.3112	0.2350	1.4457
LSTM	0.1394	0.1105	3.0073	0.2677	0.1979	1.2353

Secara umumnya, uji kaji ini menunjukkan model pembelajaran mendalam mendapat prestasi ramalan yang lebih baik berbanding model pembelajaran mesin yang lain dengan model MLP menjadi model paling baik diikuti oleh model LSTM. Model MLP mendapat nilai RMSE, MAE dan MAPE pada kadar 0.1381, 0.1095 dan 2.9761 bagi saham Axiata Group dan 0.2514, 0.1912 dan 1.1901 bagi saham Petronas Gas. Model RFR pula mendapat prestasi ramalan paling baik berbanding model DTR dan SVR dengan nilai RMSE, MAE dan MAPE pada kadar 0.1537, 0.1248 dan 3.3466 bagi saham Axiata Group dan 0.3005, 0.2238 dan 1.3909 bagi saham Petronas Gas. Ia juga turut mengatasi prestasi model CNN bagi ramalan saham Petronas Gas. Model SVR mendapat prestasi ramalan paling rendah bagi saham Axiata berbanding kesemua model lain. Bagi saham Petronas Gas pula, model DTR mendapat prestasi ramalan paling rendah berbanding yang lain.

#### b. Saiz Tettingkap 30 Hari

Hasil penilaian daripada uji kaji yang dijalankan ke atas kesemua model dengan menggunakan saiz tettingkap sebanyak 30 hari adalah seperti di dalam Jadual 4.8 di bawah:

Jadual 4.8 Analisis prestasi ramalan harga saham 5 hari ke depan dengan saiz tettingkap 30 hari

Algoritma	Axiata Group Berhad			Petronas Gas Berhad		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
DTR	0.2681	0.2156	5.7463	0.4230	0.3366	2.0997
RFR	0.1539	0.1253	3.3772	0.2923	0.2160	1.3420

bersambung...

...sambungan						
SVR	0.2118	0.1764	4.7813	0.4437	0.3686	2.2942
MLP	0.1455	0.1171	3.1565	0.2586	0.1941	1.2132
CNN	0.1341	0.1063	2.8760	0.2719	0.2093	1.3077
LSTM	0.1441	0.1140	3.0955	0.2625	0.1973	1.2321

Uji kaji menunjukkan kesemua model pembelajaran mendalam MLP, CNN dan LSTM menunjukkan prestasi ramalan paling baik berbanding model pembelajaran mesin DTR, RFR dan SVR. Model MLP mendapat prestasi ramalan paling baik dengan nilai RMSE, MAE dan MAPE pada kadar 0.1455, 0.1171 dan 3.1565 bagi saham Axiata Group dan 0.2586, 0.1941 dan 1.2132 bagi saham Petronas Gas. Model RFR pula mendapat prestasi ramalan paling baik berbanding model pembelajaran mesin DTR dan SVR dengan nilai RMSE, MAE dan MAPE pada kadar 0.1539, 0.1253 dan 3.3772 bagi saham Axiata Group dan 0.2923, 0.2160 dan 1.3420 bagi saham Petronas Gas. Prestasi ramalan paling rendah bagi saham Axiata Group diperolehi oleh model DTR manakala bagi saham Petronas Gas pula diperolehi oleh model SVR. Secara keseluruhannya, kesemua model menunjukkan peningkatan nilai RMSE, MAE dan MAPE berbanding uji kaji sebelum. Ini menunjukkan berlakunya peningkatan ralat ramalan ke atas kesemua model berbanding uji kaji sebelum ini.

### c. Saiz Tetingkap 60 Hari

Hasil penilaian daripada uji kaji yang dijalankan ke atas kesemua model dengan menggunakan saiz tetingkap sebanyak 60 hari adalah seperti di dalam Jadual 4.9 di bawah:

Jadual 4.9 Analisis prestasi ramalan harga saham 5 hari ke depan dengan saiz tetingkap 60 hari

Algoritma	Axiata Group Berhad			Petronas Gas Berhad		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
DTR	0.3404	0.2534	6.7943	0.4811	0.3740	2.3367
RFR	0.1378	0.1078	2.9201	0.2848	0.2197	1.3673
SVR	0.2125	0.1719	4.6205	0.6267	0.4917	3.0447
MLP	0.1477	0.1184	3.1728	0.3123	0.2423	1.5007
CNN	0.1379	0.1092	2.9601	0.2931	0.2230	1.3926
LSTM	0.1334	0.1054	2.8229	0.2863	0.2202	1.3765



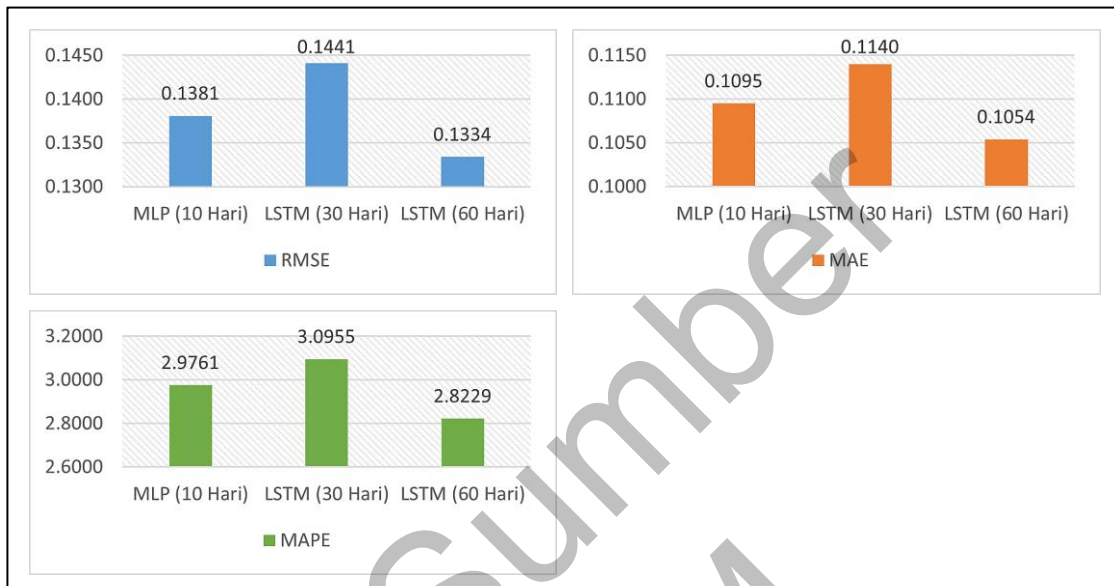
Uji kaji ini menunjukkan model LSTM mendapat prestasi ramalan paling baik berbanding kesemua model lain dengan nilai RMSE, MAE dan MAPE pada kadar 0.1334, 0.1054 dan 2.8229 bagi saham Axiata Group dan 0.2863, 0.2202 dan 1.3765 bagi saham Petronas Gas. Model RFR pula mendapat prestasi ramalan paling baik di kalangan model pembelajaran mesin dengan nilai RMSE, MAE dan MAPE pada kadar 0.1378, 0.1078 dan 2.9201 bagi saham Axiata Group dan 0.2848, 0.2197 dan 1.3673 bagi saham Petronas Gas. Ia juga turut mengatasi prestasi ramalan model pembelajaran mendalam MLP dan CNN bagi kedua-dua saham. Prestasi ramalan paling rendah bagi saham Axiata Group diperolehi oleh model DTR manakala bagi saham Petronas Gas pula diperolehi oleh model SVR.

#### **d. Analisis Prestasi Model Ramalan**

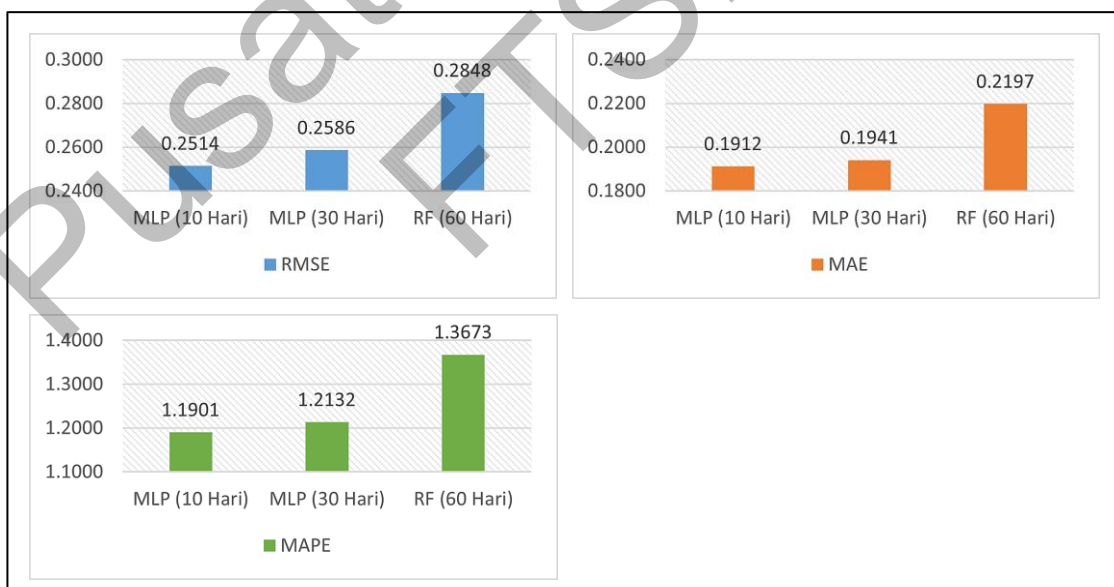
Secara keseluruhannya, model MLP dan LSTM mendapat prestasi ramalan paling baik bagi beberapa tetapan saiz tettingkap dan saham. Model RFR turut memperolehi prestasi ramalan paling baik bagi saham Petronas Gas dengan tetapan saiz tettingkap sebanyak 60 hari. Namun prestasi terbaik model RFR tersebut masih gagal mengatasi prestasi keseluruhan model MLP dan LSTM menggunakan tetapan saiz tettingkap yang berbeza. Ini dibuktikan melalui nilai RMSE, MAE dan MAPE secara keseluruhan bagi ramalan ini di mana model pembelajaran mendalam khususnya MLP dan LSTM masih mengatasi keupayaan model-model yang lain dalam membuat ramalan.

Seperti ramalan jangka panjang sebelum ini bagi tempoh tiga hari ke depan, saiz tettingkap turut mempengaruhi dan memberikan kesan yang berbeza ke atas prestasi ramalan setiap model. Terdapat model yang mengalami penurunan prestasi apabila saiz tettingkap meningkat manakala sebahagian yang lain pula mengalami peningkatan prestasi ramalan. Bagi ramalan harga saham Axiata Group, prestasi terbaik bagi saiz tettingkap 10 hari diperolehi oleh model MLP manakala saiz tettingkap 30 dan 60 hari diperolehi oleh model LSTM. Bagi saham Petronas Gas pula, prestasi ramalan terbaik bagi saiz tettingkap 10 dan 30 hari diperolehi oleh model MLP manakala model RFR untuk saiz tettingkap 60 hari. Rajah 4.9 di bawah menunjukkan perbezaan nilai RMSE, MAE dan MAPE bagi model-model yang mendapat prestasi paling baik untuk saham Axiata Group. Daripada rajah tersebut jelas menunjukkan prestasi ramalan terbaik diperolehi pada saiz tettingkap sebanyak 60 hari. Rajah 4.10 pula menunjukkan

perbezaan nilai RMSE, MAE dan MAPE bagi model-model yang mendapat prestasi paling baik untuk saham Petronas Gas. Bagi saham Petronas Gas pula, prestasi ramalan terbaik diperolehi pada saiz tettingkap sebanyak 10 hari.



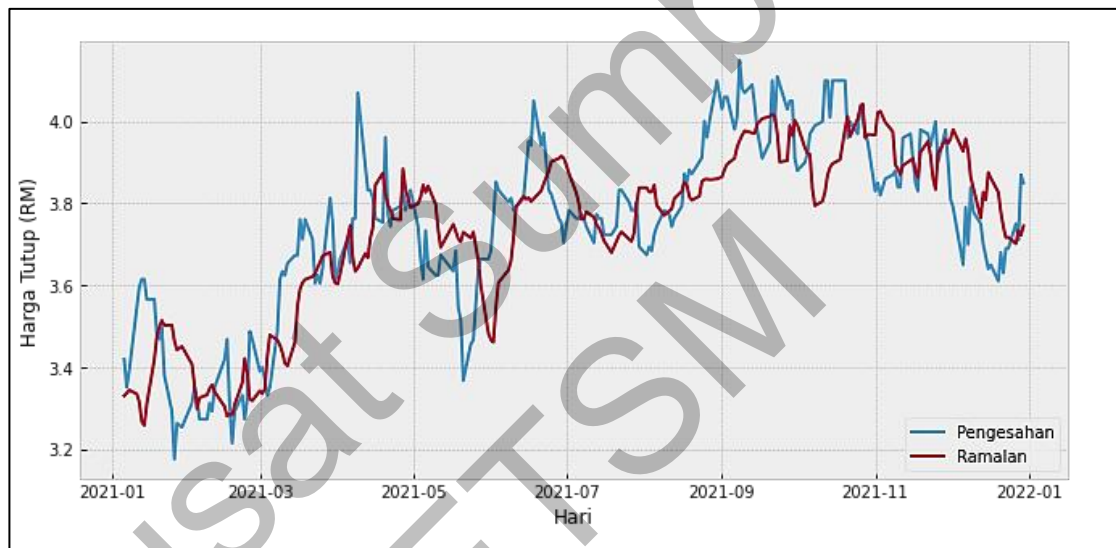
Rajah 4.9 Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Axiata Group



Rajah 4.10 Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Petronas Gas

Hasil daripada analisis ini, LSTM dengan saiz tettingkap sebanyak 60 hari adalah model terbaik untuk meramal harga saham Axiata Group bagi tempoh lima hari ke

depan manakala bagi saham Petronas Gas pula, model MLP dengan saiz tettingkap sebanyak 10 hari merupakan model yang terbaik. Kedua-dua model ini dipilih berdasarkan kepada nilai RMSE, MAE dan MAPE yang paling kecil berbanding dengan kesemua model dan saiz tettingkap yang lain. Perbandingan hasil ramalan kedua-dua model ini dengan harga tutup saham yang sebenar adalah seperti di dalam Rajah 4.11 dan Rajah 4.12 di bawah. Hasil ramalan bagi kedua-dua saham tersebut dilihat mula menunjukkan jurang perbezaan yang lebih ketara berbanding ramalan bagi tempoh tiga hari ke depan yang dibuat sebelum ini. Ini menunjukkan ralat ramalan akan menjadi semakin besar disebabkan oleh tempoh ramalan yang semakin menjauh.



Rajah 4.11 Ramalan model LSTM dengan pendekatan univariat ke atas harga tutup saham Axiata Group untuk tempoh lima hari ke depan berbanding harga sebenar



Rajah 4.12 Ramalan model MLP dengan pendekatan univariat ke atas harga tutup saham Petronas Gas untuk tempoh lima hari ke depan berbanding harga sebenar

#### 4.2.4 Ramalan Harga Saham Tujuh Hari Ke Depan

##### a. Saiz Tetingkap 10 Hari

Hasil penilaian daripada uji kaji yang dijalankan ke atas kesemua model dengan menggunakan saiz tetingkap sebanyak 10 hari adalah seperti di dalam Jadual 4.10 di bawah:

Jadual 4.10 Analisis prestasi ramalan harga saham 7 hari ke depan dengan saiz tetingkap 10 hari

Algoritma	Axiata Group Berhad			Petronas Gas Berhad		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
DTR	0.2680	0.1967	5.2591	0.4977	0.3828	2.3796
RFR	0.1660	0.1300	3.4638	0.3342	0.2519	1.5655
SVR	0.3686	0.3065	8.4270	0.4007	0.3295	2.0746
MLP	0.1645	0.1312	3.4834	0.3312	0.2551	1.5816
CNN	0.1405	0.1113	3.0039	0.3168	0.2424	1.5053
LSTM	0.1396	0.1110	2.9799	0.3122	0.2384	1.4826

Uji kaji ini menunjukkan model LSTM menjadi model yang mendapat prestasi ramalan paling baik berbanding kesemua model yang lain bagi kedua-dua saham. Ia mendapat nilai RMSE, MAE dan MAPE pada kadar 0.1396, 0.1110 dan 2.9799 bagi

saham Axiata Group dan 0.3122, 0.2384 dan 1.4826 bagi saham Petronas Gas. Model RFR pula mendapat prestasi ramalan paling baik di kalangan model pembelajaran mesin dengan nilai RMSE, MAE dan MAPE pada kadar 0.1660, 0.1300 dan 3.4638 bagi saham Axiata Group dan 0.3342, 0.2519 dan 1.5655 bagi saham Petronas Gas. Ia turut mengatasi prestasi ramalan model pembelajaran mendalam MLP bagi saham Petronas Gas. Prestasi ramalan paling rendah bagi saham Axiata Group diperolehi oleh model SVR manakala bagi saham Petronas Gas pula diperolehi oleh model DTR.

#### b. Saiz Tetingkap 30 Hari

Hasil penilaian daripada uji kaji yang dijalankan ke atas kesemua model dengan menggunakan saiz tetingkap sebanyak 30 hari adalah seperti di dalam Jadual 4.11 di bawah:

Jadual 4.11 Analisis prestasi ramalan harga saham 7 hari ke depan dengan saiz tetingkap 30 hari

Algoritma	Axiata Group Berhad			Petronas Gas Berhad		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
DTR	0.3049	0.2441	6.4732	0.4621	0.3664	2.2827
RFR	0.1627	0.1301	3.4795	0.3378	0.2497	1.5484
SVR	0.2642	0.2256	6.0886	0.5215	0.4292	2.6764
MLP	0.1538	0.1215	3.2344	0.3053	0.2312	1.4440
CNN	0.1409	0.1122	3.0074	0.3307	0.2644	1.6628
LSTM	0.1399	0.1121	3.0063	0.3193	0.2386	1.4799

Uji kaji ini menunjukkan prestasi ramalan model pembelajaran mendalam MLP, CNN dan LSTM mengatasi ketiga-tiga model pembelajaran mesin DTR, RFR dan SVR bagi kedua-dua saham. Model LSTM mendapat prestasi ramalan paling baik bagi saham Axiata Group dengan nilai RMSE, MAE dan MAPE pada kadar 0.1399, 0.1121 dan 3.0063. Model MLP pula mendapat prestasi ramalan paling baik bagi saham Petronas Gas dengan nilai RMSE, MAE dan MAPE pada kadar 0.3053, 0.2312 dan 1.4440. Model RFR mendapat prestasi ramalan paling baik di kalangan model pembelajaran mesin dengan nilai RMSE, MAE dan MAPE pada kadar 0.1627, 0.1301 dan 3.4795 bagi saham Axiata Group dan 0.3378, 0.2497 dan 1.5848 bagi saham Petronas Gas.

Prestasi ramalan paling rendah bagi saham Axiata Group diperolehi oleh model DTR manakala bagi saham Petronas Gas pula diperolehi oleh model SVR.

**c. Saiz Tetingkap 60 Hari**

Hasil penilaian daripada uji kaji yang dijalankan ke atas kesemua model dengan menggunakan saiz tetingkap sebanyak 60 hari adalah seperti di dalam Jadual 4.12 di bawah:

Jadual 4.12 Analisis prestasi ramalan harga saham 7 hari ke depan dengan saiz tetingkap 60 hari

Algoritma	Axiata Group Berhad			Petronas Gas Berhad		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
DTR	0.3733	0.2748	7.3388	0.5227	0.4213	2.6340
RFR	0.1450	0.1117	3.0041	0.3276	0.2567	1.5965
SVR	0.2504	0.2020	5.4136	0.6930	0.5560	3.4518
MLP	0.2127	0.1788	4.7816	0.3842	0.2979	1.8642
CNN	0.1467	0.1151	3.0853	0.3435	0.2584	1.6054
LSTM	0.1540	0.1266	3.3916	0.3577	0.2780	1.7337

Uji kaji ini pula menunjukkan model pembelajaran mesin RFR mendapat prestasi ramalan paling baik berbanding kesemua model lain dan mengatasi prestasi ketiga-tiga model pembelajaran mendalam. Ia juga menunjukkan peningkatan prestasi ramalan berbanding uji kaji sebelum ini dengan nilai RMSE, MAE dan MAPE pada kadar 0.1450, 0.1117 dan 3.0041 bagi saham Axiata Group dan 0.3276, 0.2567 dan 1.5965 bagi saham Petronas Gas. Model CNN pula mendapat prestasi ramalan paling baik di antara semua model pembelajaran mendalam dengan nilai RMSE, MAE dan MAPE pada kadar 0.1467, 0.1151 dan 3.0853 bagi saham Axiata Group dan 0.3435, 0.2584 dan 1.6054 bagi saham Petronas Gas. Prestasi ramalan paling rendah bagi saham Axiata Group diperolehi oleh model DTR manakala bagi saham Petronas Gas pula diperolehi oleh model SVR. Secara keseluruhannya, kesemua model menunjukkan peningkatan ralat ramalan berbanding uji kaji sebelum ini kecuali bagi model RFR dan SVR di mana keduanya menunjukkan penurunan nilai RMSE, MAE dan MAPE yang ketara berbanding sebelumnya.

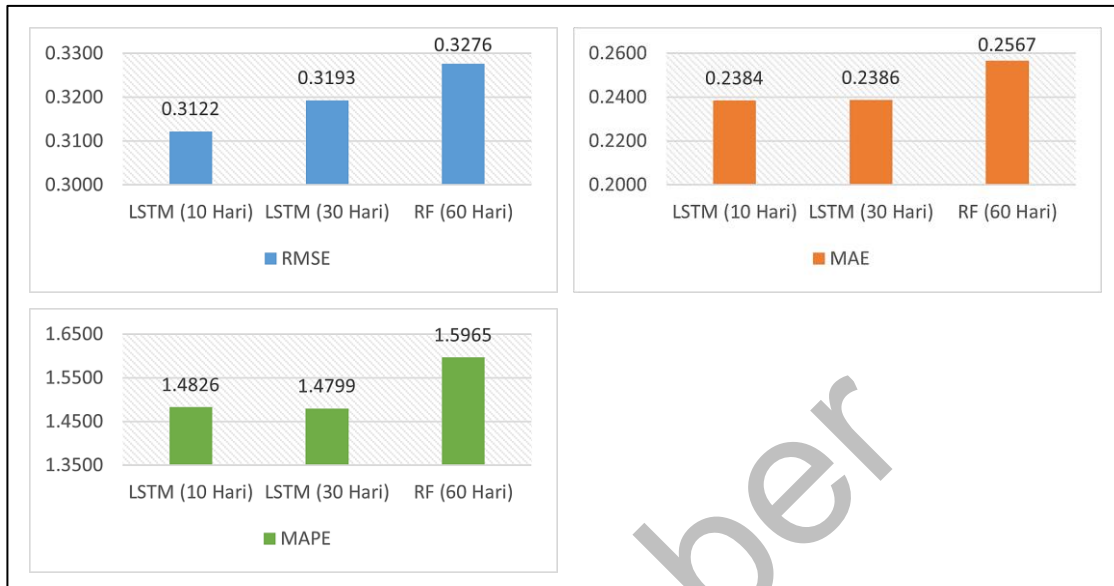
#### d. Analisis Prestasi Model Ramalan

Kesemua model pembelajaran mendalam dan model DTR dilihat mengalami penurunan prestasi apabila saiz tettingkap meningkat. Namun bagi model RFR dan SVR, prestasinya menunjukkan peningkatan apabila saiz tettingkap membesar. Prestasi keseluruhan bagi ramalan untuk tempoh tujuh hari ke depan ini masih diperolehi oleh ketiga-tiga model pembelajaran mendalam MLP, CNN dan LSTM pada tetapan saiz tettingkap yang berbeza. Model RFR turut mendapat prestasi terbaiknya pada saiz tettingkap 60 hari bagi kedua-dua saham.

Untuk kedua-dua saham Axiata Group dan Petronas Gas, prestasi terbaik ramalan menggunakan saiz tettingkap sebanyak 10 dan 30 hari dicapai oleh model LSTM manakala saiz 60 hari dicapai oleh model RFR. Sebagai perbandingan, Rajah 4.13 dan Rajah 4.14 di bawah menunjukkan perbezaan nilai RMSE, MAE dan MAPE yang diperolehi oleh setiap model tersebut mengikut saiz tettingkap yang ditetapkan untuk saham Axiata Group dan Petronas Gas. Berdasarkan rajah tersebut, pemilihan saiz tettingkap sebanyak 10 hari menghasilkan prestasi ramalan paling baik bagi kedua-dua saham berbanding saiz tettingkap yang lain.



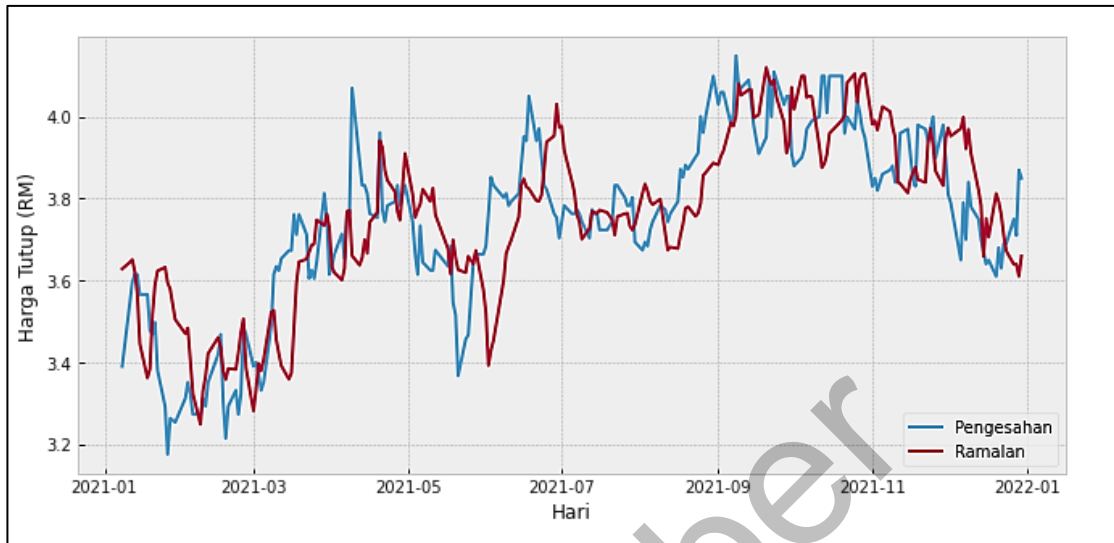
Rajah 4.13 Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Axiata Group



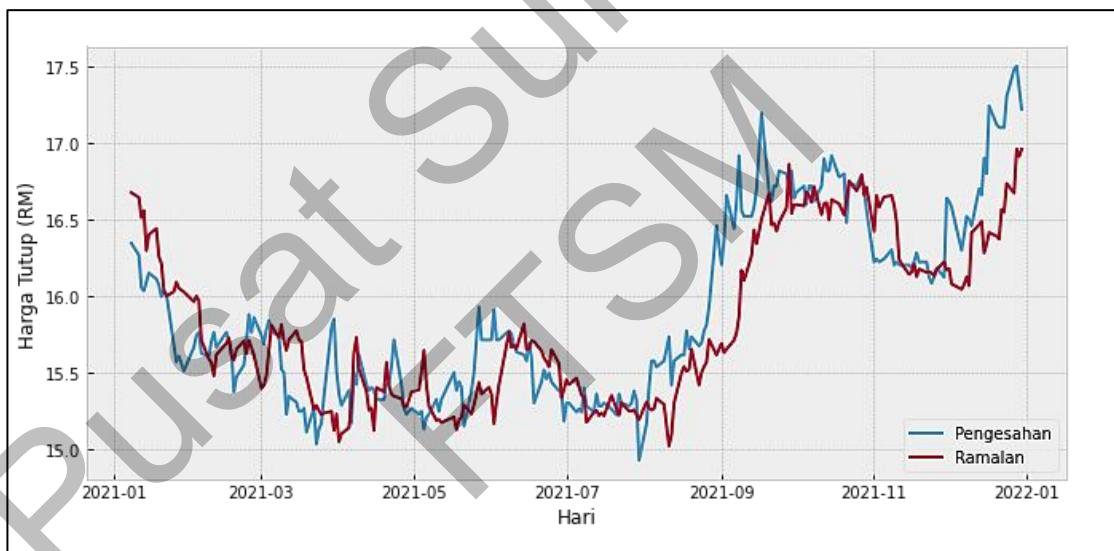
Rajah 4.14 Perbandingan nilai RMSE, MAE dan MAPE model terbaik untuk setiap saiz tettingkap bagi saham Petronas Gas

Berdasarkan analisis ini, model LSTM dengan tetapan saiz tettingkap sebanyak 10 hari adalah model terbaik bagi meramal harga tutup kedua-dua saham untuk tempoh tujuh hari ke depan. Ini dibuktikan dengan nilai RMSE, MAE dan MAPE yang diperolehi oleh model ini adalah paling kecil berbanding dengan kesemua model dan saiz tettingkap yang lain. Perbandingan hasil ramalan model ini bagi kedua-dua saham dengan harga tutup saham yang sebenar adalah seperti di dalam Rajah 4.15 dan Rajah 4.16 di bawah. Hasil ramalan bagi kedua-dua saham tersebut turut menunjukkan jurang perbezaan yang agak ketara seperti ramalan bagi tempoh tiga dan lima hari ke depan yang dibuat sebelum ini. Namun, ralat ramalan tersebut masih boleh diterima di mana ianya berada di dalam jurang yang tidak terlalu menjauh dari nilainya yang sebenar.





Rajah 4.15 Ramalan model MLP dengan pendekatan univariat ke atas harga tutup saham Axiata Group untuk tempoh tujuh hari ke depan berbanding harga sebenar



Rajah 4.16 Ramalan model MLP dengan pendekatan univariat ke atas harga tutup saham Petronas Gas untuk tempoh tujuh hari ke depan berbanding harga sebenar

### 4.3 RAMALAN MULTIVARIAT

Pendekatan multivariat pula dalam membuat ramalan harga saham adalah kaedah yang berasaskan kepada pemerhatian pelbagai fitur daripada set data yang digunakan. Kesemua fitur tersebut adalah daripada tempoh dan jujukan masa yang sama. Bagi data kajian ini, fitur-fitur yang digunakan adalah harga buka, harga tutup, harga terendah, harga tertinggi, dan jumlah dagangan. Pendekatan ini digunakan ke atas kedua-dua data