

PEMBANGUNAN APLIKASI TEKNIK PEMBELAJARAN MESIN DALAM RAMALAN PASARAN SAHAM

Ng Wei Sheng

Shahrul Azman Mohd Noah

*Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia, 43600 UKM Bangi,,
Selangor Darul Ehsan, Malaysia*

Abstrak

Pasaran saham telah lama menjadi cara bagi individu untuk menghasilkan pendapatan tambahan, di mana para pelabur dapat mendapatkan keuntungan daripada perbezaan harga dan pembayaran dividen. Pelabur terkenal, Warren Buffet, merupakan contoh utama individu yang telah mencapai kejayaan besar di pasaran saham. Dengan kemajuan teknologi kecerdasan buatan, kajian mengenai ramalan pasaran saham telah menjadi popular dalam menentukan potensi pembelajaran mesin untuk meramalkan pergerakan harga saham pada masa depan. Dalam kajian ini, tiga algoritma pembelajaran mesin, iaitu *linear regression*, *multilayer perceptron (MLP)*, dan juga *long short term memory (LSTM)* telah dipakai. Selain itu, teknik penggabungan (*bagging*) juga telah diperkenalkan untuk meningkatkan prestasi model *LSTM*. Dataset sejarah Maxis dari 2/1/2018 hingga 30/3/2023 telah digunakan dalam kajian ini. Beberapa langkah pra-pemrosesan, seperti menangani nilai yang hilang, menghapuskan nilai pencilan (*outliers*), dan menukar jenis data telah dilakukan untuk memastikan kesesuaian data untuk pemodelan. Selain itu, dua ciri tambahan, iaitu purata pergerakan saham 14 hari dan 26 hari, telah diperkenalkan. Analisis perbandingan ketiga-tiga algoritma ini menunjukkan bahawa hanya model *LSTM* yang dapat meramalkan harga saham, dengan nilai R^2 sebanyak 0.7972. Sebaliknya, model *linear regression* menghasilkan nilai R^2 sebanyak -4.7906, dan model *MLP* mempunyai nilai R^2 sebanyak -40.7682.

Pengenalan

Menurut kepada Dictionary.com, pasaran saham dimaksudkan sebagai pasar di mana saham diperdagangkan. Ketika individu membeli saham perusahaan publik, mereka pada dasarnya membeli sebagian dari perusahaan tersebut. Harga saham dapat naik atau turun tergantung pada permintaan dan penawaran saham (*demand and supply*). Secara umum, ketika lebih banyak orang membeli saham, harga naik, dan ketika lebih banyak orang menjual saham, harga turun.

Terdapat banyak teknik untuk mendapatkan keuntungan di pasar saham seperti mampu membaca tren, mengenal pasti bahawa syarikat tersebut mempunyai bisnes yang baik, melabur ketika saham kurang nilai (*undervalued*) dan sedang mempunyai aliran menaik (*uptrend*), dan sebagainya. Namun, penyelidikan menunjukkan bahawa sebanyak 90% orang mencatat kehilangan wang di pasar saham. Banyak faktor yang menyebabkan kegagalan tersebut. Salah satu faktor yang paling umum adalah kekurangan pengetahuan tentang pasar saham. Untuk memahami pasar saham, individu perlu memiliki keupayaan dalam membaca penyataimbangan (*balance sheet*) dan pernyataan pendapatan (*income statement*) yang dikeluarkan oleh syarikat. Namun, hal ini memerlukan ilmu pengetahuan dalam kewangan dan perakaunan, yang boleh menjadikan kelemahan bagi orang di luar bidang tersebut. Selain itu, komitmen harian dalam pekerjaan atau pembelajaran juga telah menyekat individu untuk sentiasa membaca berita terbaru, yang menyebabkan ketidakmampuan untuk bereaksi terhadap pasar saham.

Dalam projek ini, tiga algoritma pembelajaran mesin iaitu *linear regression*, *MLP*, dan *LSTM* telah digunakan untuk membangun peramal harga saham. Tujuan projek ini adalah untuk menilai kemampuan teknologi kecerdasan buatan dalam meramal harga saham, dan juga membantu pelabur dalam membuat keputusan pelaburan yang terinformasi dan menghindari daripada potensi risiko. Projek ini akan menggunakan dataset sejarah harga saham Maxis untuk tujuan penyelidikan. Dataset tersebut mengandungi ciri-ciri seperti *Date*, *Open*, *High*, *Low*, *Close*, *Adj Close*, dan *Volume*. Penjelasan ciri-ciri adalah sebagai berikut:

1. *Date*

Tarikh spesifik di mana harga saham dicatatkan.

2. *Open*

Harga di mana saham mulai diperdagangkan pada awal hari perdagangan.

3. *High*

Harga tertinggi yang tercatat dalam satu hari perdagangan.

4. *Low*

Harga terendah yang tercatat dalam satu hari perdagangan.

5. *Close*

Harga terakhir di mana saham diperdagangkan pada akhir hari perdagangan.

6. *Adj Close*

Harga penutupan saham yang telah diubahsuai bagi mempertimbangkan tindakan perusahaan yang dapat mempengaruhi nilai saham.

7. *Volume*

Jumlah saham yang diperdagangkan dalam satu hari perdagangan.

Selain itu, dua ciri tambahan, iaitu purata pergerakan saham dalam 14 hari (*sma_14/14-days simple moving averages*) dan purata pergerakan saham dalam 26 hari (*sma_26 /26-days simple moving averages*), telah diperkenalkan dalam penyelidikan ini, menjadikan jumlah bilangan ciri kepada sembilan. Purata pergerakan saham mengira nilai purata pergerakan harga

penutupan selama jumlah hari yang ditentukan. Misalnya, *sma_14* bermaksud purata harga selama 14 hari, dan *sma_26* bermaksud purata harga selama 26 hari. Dalam konteks pasar saham, ciri-ciri ini biasanya dipakai dalam analisis teknik pergerakan harga saham. Secara umum, ketika purata bergerak yang lebih tinggi (*sma_26*) berada di bawah purata bergerak yang lebih rendah (*sma_14*) dan keduanya bergerak ke atas, itu menunjukkan aliran menaik dalam harga saham yang memberikan potensi untuk melabur. Di samping itu, jika purata bergerak yang lebih tinggi (*sma_26*) berada di atas rata-rata bergerak yang lebih rendah (*sma_14*) dan keduanya bergerak ke bawah, itu menunjukkan aliran menurun dalam harga saham.

8. *sma_14*

Purata bergerak harga penutupan selama 14 hari.

9. *sma_26*

Purata bergerak harga penutupan selama 26 hari.

Terdapat lima objektif dalam projek ini, iaitu:

- a. Membina peramal harga saham menggunakan algoritma pembelajaran mesin.
- b. Menentukan keupayaan pembelajaran mesin dalam meramalkan harga saham.
- c. Mengenal pasti algoritma pembelajaran mesin (*linear regression*, *MLP* atau *LSTM*) yang paling sesuai untuk sistem ramalan harga saham.
- d. Mengenal pasti ciri-ciri yang terbaik untuk menghasilkan ramalan yang berkesan.
- e. Menunjukkan harga saham yang diramalkan dalam laman web.

Sehingga Julai 2022, terdapat kira-kira 982 syarikat yang disenaraikan di Bursa Malaysia.

Bursa Malaysia merupakan bursa saham di Malaysia. Bursa Malaysia telah mengategorikan

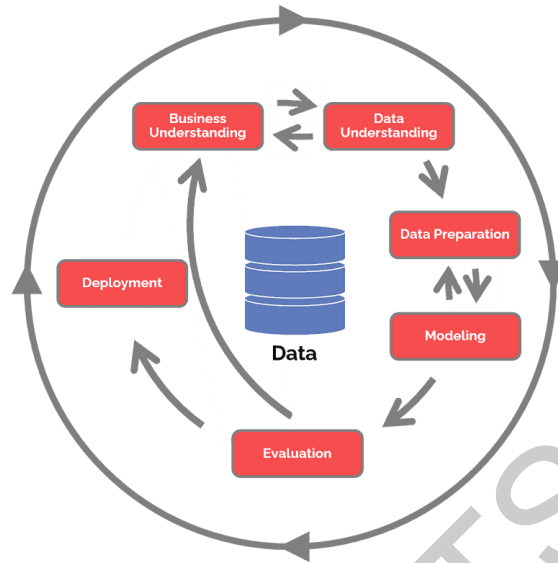
syarikat-syarikat tersebut ke dalam beberapa sektor. Lima sektor utama di Bursa Malaysia adalah perkhidmatan kewangan, produk dan perkhidmatan pengguna, produk dan perkhidmatan industri, telekomunikasi dan media, dan juga penjagaan kesihatan. Dalam projek ini, fokus akan diberikan kepada sektor telekomunikasi dan media. Antara syarikat-syarikat yang dikandungi dalam sektor ini adalah Digi, Maxis, Axiata, dan lain-lain.

Tumpuan utama sistem adalah visualisasi pergerakan pasaran saham yang diramalkan menggunakan graf garis. Selain itu, sistem ini juga akan memaparkan maklumat pasaran saham semasa. Untuk memastikan pengguna sentiasa mendapat berita terkini, artikel-artikel berkaitan dengan syarikat-syarikat telekomunikasi juga akan diterbitkan dalam sistem tersebut.

Secara keseluruhan, sistem ini akan mengandungi:

- a. Ramalan harga saham bagi syarikat-syarikat yang disenaraikan di Bursa Malaysia dalam sektor telekomunikasi dan media (Maxis, Digi, Axiata, Celcom, TimeDotCom).
- b. Artikel-artikel mengenai syarikat-syarikat yang disenaraikan di Bursa Malaysia dalam sektor telekomunikasi dan media.

Projek ini akan memakai metodologi *CRISP-DM*, yang merupakan singkatan dari "*Cross-Industry Standard Process for Data Mining*". Ini adalah kitaran hidup projek yang biasanya digunakan untuk menangani masalah perlombongan data. *CRISP-DM* terdiri daripada enam fasa: fasa pemahaman bisnes, fasa pemahaman data, fasa penyediaan data, fasa pemodelan, fasa penilaian, dan fasa penyebaran. Rajah 1.1 menunjukkan rangka kerja *CRISP-DM*.

Rajah 1 Rangka kerja *CRISP-DM*

Sumber: <https://www.datascience-pm.com/crisp-dm-2/>

Baki kertas ini disusun sebagai berikut: Metodologi kajian menjelaskan proses CRISP-DM secara terperinci, keputusan dan perbincangan menunjukkan penilaian model dan pengujian sistem dalam kajian ini, diikuti oleh kesimpulan yang menghuraikan rumusan dari kertas ini dan juga cadangan untuk penyelidikan masa depan.

Sorotan Susastera

Dalam kertas *Stock Price Prediction Using Machine Learning* oleh Yixin Guo, penulis menggunakan data sejarah S&P 500 untuk penyelidikannya. Penulis memperkenalkan tiga algoritma, iaitu *ARIMA*, *LSTM*, dan *GARCH*. Selain itu, penulis juga menggunakan pembelajaran ensemble dengan menggabungkan algoritma-algoritma tersebut. Namun, kertas tersebut tidak memberi perincian tentang teknik pra-pemprosesan yang digunakan dan ciri-ciri spesifik yang telah dipakai untuk latihan model. Hasil penyelidikan menunjukkan bahawa ensemble dari algoritma-algoritma tersebut mencapai *MSE* terendah (0.412), diikuti oleh *LSTM* (0.876), *ARIMA* (1.213), dan *GARCH* (1.923).

Dalam kertas *Stock Price Prediction Using Machine Learning Techniques* oleh Jagruti Hota, Sujata Chakravarty, Bijay K. Paikaray dan Harshvardhan Bhojar, penulis menggunakan data sejarah American Airlines untuk menyelidikannya. Penulis memperkenalkan empat algoritma, yaitu *decision tree (DT)*, *support vector regressor (SVR)*, *random forest (RF)*, dan *artificial neural network (ANN)*. Penulis menggunakan beberapa teknik preprocessing termasuk imputasi nilai hilang dengan nilai purata, penskalaan data, dan normalisasi. Data yang telah dibersihkan kemudian dibagi menjadi set latihan dan ujian dalam nisbah 7:3. Ciri-ciri yang digunakan untuk perahan ciri (*feature extraction*) adalah harga pembukaan atau kolom 'open' saham American Airlines. Hasil penelitian menunjukkan bahwa *RF* mencapai *mean absolute percentage error* terendah (0.36), diikuti oleh *ANN* (0.37), *DT* (1.60), dan *SVR* (3.56).

Dalam kertas *Stock Price Predictions Using Machine Learning Algorithms* oleh Shreya Pawaskar, penulis menggunakan data sejarah Tata Consultancy Services untuk menyelidikannya. Penulis memperkenalkan empat algoritma, yaitu *multiple linear regression (MLR)*, *polynomial regression (PR)*, *decision tree regressor (DTR)*, dan *random forest regressor (RFR)*. Penulis memperkenalkan dua ciri tambahan yaitu '*Day_Perc_Change*' yang menunjukkan peratusan perubahan dalam satu hari dari harga penutupan, dan '*HL_PCT*' yang menunjukkan perbezaan antara harga tertinggi dan terendah dibagi dengan *Adj Close Price*. Menggunakan kepentingan ciri, kolom-kolom *Day_Perc_Change*, *HL_PCT*, *Adj Close Price*, dan *Volume* dipilih sebagai pembolehubah tidak bersandar. Hasil penyelidikan menunjukkan bahwa *DTR* mencapai *root mean squared error* terendah (0.0), diikuti oleh *RFR* (32.45), *PR* (102.29), dan *MLR* (106.42).

Dalam kertas *Stock Closing Price Prediction Using Machine Learning Techniques* oleh Mehar Vijn, Deeksha Chandola, Vinay Anand Tikkiwal dan Arun Kumar, penulis menggunakan data sejarah Nike, Goldman Sachs, JP Morgan and Co., Johnson & Johnson, dan Pfizer Inc. Penulis memperkenalkan dua algoritma, yaitu *ANN* dan *RF*. Penulis juga memperkenalkan tujuh

ciri tambahan iaitu perbezaan antara harga tertinggi dan terendah saham ($H-L$), perbezaan antara harga penutupan dan pembukaan saham ($O-C$), purata bergerak tujuh hari harga saham ($7 DAYS MA$), purata bergerak empat belas hari harga saham ($14 DAYS MA$), purata bergerak dua puluh satu hari harga saham ($21 DAYS MA$), dan sisihan piawai (*standard deviation*) harga saham untuk tujuh hari terakhir ($7 DAYS STD DEV$). Secara keseluruhan, ANN (rata-rata $MAPE$ 0.904%) lebih baik dalam meramalkan harga saham syarikat-syarikat tersebut dibandingkan dengan RF (rata-rata $MAPE$ 0.926%).

Dalam kertas *Research on the Prediction Method of Stock Price Based on RBF Neural Network Optimization Algorithm* oleh Hongzheng Li dan Shaohang Huang, penulis tidak menyebutkan nama syarikat yang digunakan untuk tujuan penyelidikan, tetapi hanya dengan kod saham MET, CCU, dan TR. Penulis hanya memperkenalkan rangkaian neural RBF sebagai algoritma yang digunakan, dengan kombinasi algoritma *K-means clustering* untuk mengoptimalkan parameter rangkaian. Penulis tidak menyebutkan langkah-langkah pra-pemrosesan yang diambil, tetapi hasil penyelidikan menunjukkan bahwa algoritma tersebut mencapai nilai MSE rata-rata sebanyak 3.567.

Metodologi Kajian

1. Fasa pemahaman bisnes

Tujuan dan keperluan projek dikenal pasti dan diterjemahkan menjadi definisi masalah perlombongan data. Strategi awal, seperti memperoleh perkekas dan perisian yang diperlukan serta mengenal pasti sumber data, juga akan dilakukan.

2. Fasa pemahaman data

Proses pengumpulan data dilakukan menggunakan platform atas talian. *Yahoo Finance* telah menjadi sumber pengumpulan data, disebabkan platform tersebut mempunyai data sejarah pasar saham di seluruh dunia. Data dimuat turun dari *Yahoo Finance* dalam format *CSV*. Selain itu, analisis data penyelidikan (*EDA*) dilakukan untuk memahami data dan mendapatkan wawasan awal.

3. Fasa penyediaan data

Sebelum data dimasukkan ke dalam model pembelajaran mesin, dataset telah dibersihkan dan diubah menggunakan berbagai teknik pra-pemrosesan. Ini termasuk penanganan data yang hilang, penskalaan, penghapusan nilai pencilan, dan lain-lain untuk memastikan bahawa data adalah sesuai untuk fasa pemodelan serta meningkatkan kecekapan proses pelatihan. Dataset yang telah dibersihkan kemudian dibagikan kepada set latihan dan ujian dalam nisbah 8:2 untuk tujuan pelatihan dan penilaian.

4. Fasa Pemodelan

Dalam fasa ini, 3 algoritma pembelajaran mesin iaitu *linear regression*, *multilayer perceptron (MLP)* dan *Long Short Term Memory (LSTM)* telah dipilih dan dibangunkan sebagai model. Hiperparameter model juga ditentukan, dan set latihan digunakan untuk melatih model. Selain itu, penalaan hiperparameter (*hyperparameter tuning*) telah dilakukan untuk mencari tetapan optimal bagi model-model tersebut.

5. Fasa Penilaian

Selepas melatih model-model tersebut, proses penilaian dilakukan dengan menggunakan metrik penilaian yang sesuai seperti *MSE* dan R^2 . Proses ini dilakukan dengan

menggunakan set ujian. Keberatan bagi setiap pemboleh ubah juga diperiksa untuk menentukan sama ada ciri tersebut signifikan bagi hasil output. Fasa penilaian diulang dengan menggunakan subset pemboleh ubah yang berbeza. Akhirnya, pilihan model dan ciri-ciri ditentukan berdasarkan hasil penilaian. Produk dari fasa ini adalah model yang berprestasi terbaik serta ciri-ciri yang paling berguna untuk latihan model.

6. Fasa Penyebaran

Ramalan harga masa depan telah dibuat menggunakan model yang paling berprestasi. Kemudian, laman web untuk menggambarkan ramalan harga saham juga dibina menggunakan bahasa pengaturcaraan seperti *HTML* dan *CSS*. Keputusan ramalan telah dimuat naik ke laman web yang dibina dalam bentuk graf garis.

Keputusan dan Perbincangan

Prestasi setiap model telah dinilai menggunakan *MSE* dan R^2 . Hasilnya menunjukkan bahawa hanya model *LSTM* yang menunjukkan ramalan berkesan harga penutup dalam dataset. Ini ditunjukkan dari nilai R^2 positif yang diperoleh untuk *LSTM* (0.7854), manakala model *linear regression* menghasilkan nilai R^2 negatif (-4.706) dan model *MLP* mempunyai nilai R^2 negatif yang lebih rendah (-32.6102). Model *linear regression* dan *MLP* dilatih semata-mata menggunakan ciri "Date", kerana ciri-ciri lain seperti "Open", "High", "Low", dan lain-lain tidak tersedia semasa fasa pengedaran apabila model perlu membuat ramalan untuk masa depan. Model *LSTM*, sebaliknya, dilatih menggunakan semua ciri yang tersedia kecuali lajur "Close", yang berfungsi sebagai ciri bersandar untuk projek ini. Tambahan pula, ciri "Adj Close" tidak digunakan dalam ketiga-tiga model kerana ciri bersandar "Close" bergantung pada ciri ini. Walaupun menghadapi cabaran data yang hilang, model *LSTM* mempunyai kelebihan dalam menganalisis data sebelumnya dan membuat ramalan berdasarkan maklumat tersebut. Ia tidak

bergantung pada rekod masa depan disebabkan keupayaannya untuk melihat ke belakang dan membuat kesimpulan tentang hasil masa depan yang berkemungkinan. Tambahan pula, model telah diselaraskan menggunakan *RandomizedSearchCV* untuk mengenal pasti parameter terbaik, menghasilkan prestasi optimum. Jadual 1 menunjukkan perbandingan antara model *linear regression*, *MLP* dan *LSTM*.

Jadual 1 Perbandingan antara model *linear regression*, *MLP* dan *LSTM*

Model	Parameter	Ciri	MSE	R ²
<i>Linear regression</i>	-	Date	0.0380	-4.7906
<i>MLP</i>	<i>activation = tanh</i> <i>batch_size = 10</i> <i>hidden_layer_sizes = (30, 20)</i> <i>learning_rate_init = 0.001</i> <i>max_iter = 20</i> <i>look_back = 25</i> <i>units = 30</i>	Date	0.2733	-40.7682
<i>LSTM</i>	<i>epoch = 15</i> <i>batch_size = 10</i> <i>activation = tanh</i>	<i>Open, High, Low, Volume,</i> <i>sma_14, sma_26</i>	0.0014	0.7854

* Nilai *MSE* di sini adalah dikira dengan nilai ternormalisasi

Seterusnya, pemilihan ciri telah dilakukan untuk mengenal pasti subset ciri yang dapat menghasilkan nilai *R²* tertinggi. Proses ini melibatkan penggunaan *recursive feature elimination (RFE)*, di mana ciri yang mempunyai keberatan atau pekali terendah akan digugur dan model diuji sekali lagi. Proses ini diulang sehingga nilai *R²* terbaik dicapai. Selepas memodelkan *LSTM*, keberatan keseluruhan setiap ciri telah dikira. Ini telah didapati bahawa *sma_26* mempunyai

berat tertinggi iaitu 14.1748, menunjukkan pengaruhnya yang kuat terhadap ramalan model. Sebaliknya, *sma_14* mempunyai berat terendah iaitu 12.4254, menunjukkan kesan yang relatif lemah terhadap prestasi model. Jadual 2 menunjukkan keberatan keseluruhan setiap ciri kepada outputnya.

Jadual 2 Keberatan keseluruhan setiap ciri

Ciri	Keberatan
<i>Open</i>	12.4315
<i>High</i>	13.3428
<i>Low</i>	13.5390
<i>Volume</i>	13.8283
<i>sma_14</i>	12.4254
<i>sma-26</i>	14.1748

Ciri *sma_14* kemudiannya digugur daripada set pembolehubah tidak bersandar. Bagaimanapun, hasil penilaian menunjukkan bahawa pengecualian *sma_14* tidak meningkatkan prestasi model. Ini dibuktikan dengan penurunan nilai R^2 daripada 0.7854 kepada 0.7553 dan peningkatan nilai MSE daripada 0.0014 kepada 0.0016. Oleh itu, ini dapat disimpulkan bahawa semua pembolehubah tidak bersandar yang dipakai dari awal harus dimasukkan ke dalam model untuk mencapai prestasi terbaik. Jadual 3 menunjukkan perbandingan prestasi model *LSTM* sebelum dan selepas menggugurkan *sma_14*.

Jadual 3 Prestasi model *LSTM* sebelum dan selepas menggugurkan *sma_14*

	<i>MSE</i>	<i>R</i> ²
Sebelum menggugur <i>sma_14</i>	0.0014	0.7854
Selepas menggugur <i>sma_14</i>	0.0016	0.7553

* Nilai *MSE* di sini adalah dikira dengan nilai ternormalisasi

Untuk meningkatkan lagi prestasi model, teknik pembelajaran ensemble yang dikenali sebagai *bagging* telah dilaksanakan dalam penyelidikan ini. *Bagging* melibatkan latihan pelbagai model *LSTM* pada subset data latihan yang berbeza menggunakan *bootstrapping* dengan penggantian. Output model-model tersebut kemudian dikumpulkan untuk mendapatkan ramalan akhir. Dalam penyelidikan ini, ensemble bagged sebanyak 10 model *LSTM* telah dibangun. Model bagged telah dinilai dan didapati bahawa *bagging* memberikan sedikit peningkatan berbanding model *LSTM* individu. Nilai *MSE* menurun daripada 0.0014 kepada 0.0013 (nilai ternormalisasi), menunjukkan peningkatan ketepatan. Tambahan pula, nilai *R*² meningkat daripada 0.7854 kepada 0.7972, menunjukkan susunan model yang lebih baik pada data. Ini menunjukkan bahawa ensemble model *LSTM* menghasilkan ramalan yang lebih tepat dan boleh dipercayai berbanding menggunakan model *LSTM* tunggal. Jadual 4 menunjukkan perbandingan prestasi model *LSTM* dengan dan tanpa *bagging*.

Jadual 4 Prestasi model *LSTM* dengan dan tanpa *bagging*

	<i>MSE</i>	<i>R</i> ²
Tanpa <i>bagging</i>	0.0014	0.7854
Dengan <i>bagging</i>	0.0013	0.7972

* Nilai *MSE* di sini adalah dikira dengan nilai ternormalisasi

Selepas memperoleh model yang berfungsi, ramalan telah dibuat untuk harga penutupan masa depan. Dalam penyelidikan ini, harga penutupan untuk 30 hari akan datang selepas baris terakhir dataset (30/3/2023) telah diramalkan. Kaedah ramalan melibatkan pengimbasan 30 rekod sebelumnya. Sebagai contoh, pada hari pertama, algoritma akan mengimbas 30 baris terakhir dataset dan membuat ramalan berdasarkananya. Pada hari kedua, ia akan mengimbas 29 baris terakhir dataset bersama-sama dengan ramalan hari pertama, dan seterusnya. Proses ini diteruskan sehingga harga untuk 30 hari telah diramalkan. Bukan seperti *linear regression* dan *MLP*, kaedah ramalan ini tidak memerlukan nilai pembolehubah tidak bersandar. Jadual 5 menunjukkan harga ramalan dan sebenar Maxis dari 31/3/2023 to 17/5/2023 (30 hari)

Jadual 5 Harga ramalan dan sebenar Maxis dari 31/3/2023 ke 17/5/2023

Tarikh	Harga ramalan	Harga sebenar
31/3/2023	4.006517	4.13
3/4/2023	4.01534	4.12
4/4/2023	4.026	4.1
5/4/2023	4.049846	4.08
6/4/2023	4.066249	4.08
7/4/2023	4.062628	4.09
10/4/2023	4.038959	4.12
11/4/2023	4.026488	4.12
12/4/2023	3.999461	4.16
13/4/2023	3.972241	4.17
14/4/2023	3.96742	4.19

17/4/2023	3.963522	4.28
18/4/2023	3.965234	4.3
19/4/2023	3.971753	4.29
20/4/2023	3.980599	4.34
25/4/2023	3.98315	4.35
26/4/2023	3.974329	4.3
27/4/2023	3.958509	4.39
28/4/2023	3.923363	4.4
2/5/2023	3.894899	4.43
3/5/2023	3.869708	4.52
5/5/2023	3.858346	4.59
8/5/2023	3.859158	4.37
9/5/2023	3.873089	4.34
10/5/2023	3.89477	4.42
11/5/2023	3.918402	4.4
12/5/2023	3.933531	4.33
15/5/2023	3.946491	4.37
16/5/2023	3.949574	4.38
17/5/2023	3.953893	4.35

Nilai MSE and R^2 antara harga ramalan dan harga sebenar telah dikira. Jadual 6 menunjukkan bahawa nilai MSE yang didapati adalah 0.1390, menunjukkan kesalahan yang agak

tinggi, dan nilai R^2 adalah -6.3766, jauh lebih rendah daripada semasa fasa ujian. Ini menunjukkan bahawa model *LSTM* boleh membuat ramalan harga saham hanya apabila data tersedia.

Jadual 6 Nilai *MSE* and R^2 antara harga ramalan dan harga sebenar

<i>MSE</i>	R^2
0.1390	-6.3766

Kesimpulan

Penglibatan semua ciri kecuali *Date* dan *Adj Close*, termasuk *sma_14* dan *sma_26* yang diperoleh daripada ciri *Close* dalam model *LSTM* telah menunjukkan peningkatan dalam prestasinya. Guguran *sma_14* telah menyebabkan nilai R^2 menurun dari 0.7854 kepada 0.7553, menunjukkan kepentingan ciri ini dalam ramalan harga saham. Penerapan pembelajaran ensemble melalui *bagging* selanjutnya meningkatkan prestasi model *LSTM*, dengan peningkatan nilai R^2 kepada 0.7972. Ini menunjukkan keberkesanan teknik ensemble dalam meningkatkan prestasi model. Walau bagaimanapun, walaupun *LSTM* menunjukkan prestasi yang baik dalam meramalkan harga saham dengan nilai R^2 yang agak tinggi, ia masih tidak dapat meramalkan harga saham masa depan dengan tepat apabila data adalah tidak didedah (*unforseen*). Ini menunjukkan bahawa *LSTM* hanya mampu meramalkan harga saham berdasarkan data yang ada, dan bukannya memberikan harga masa depan yang tepat. Oleh itu, ini dapat disimpulkan bahawa *LSTM* adalah model yang sesuai untuk meramalkan harga saham berdasarkan data sejarah, tetapi kelemahannya harus diakui ketika cuba untuk meramalkan harga saham masa depan.

Sesungguhnya, harga saham dipengaruhi oleh pelbagai faktor selain harga sejarah. Faktor luaran seperti peristiwa politik, keadaan sosio-ekonomi, prestasi syarikat, dan juga berita boleh memberi impak yang ketara terhadap harga saham. Contohnya, kejatuhan harga saham AirAsia semasa tempoh PKP COVID-19 menunjukkan bagaimana sekatan perjalanan menjejaskan pendapatan syarikat. Sebaliknya, pengeluar sarung tangan seperti Top Glove dan Hartalega menunjukkan kenaikan harga saham disebabkan peningkatan keperluan sarung tangan semasa pandemik. Contoh-contoh ini menunjukkan kerumitan ramalan harga saham dan keperluannya untuk mengambil kira pelbagai faktor dan bukan sahaja data sejarah.

Dalam penyelidikan masa depan, pengintegrasian pemrosesan bahasa tabii (*NLP*) untuk menganalisis sentimen yang terdapat dalam artikel atau berita boleh bermanfaat. Dengan mengambil kira sentimen pelabur yang memainkan peranan penting dalam mempengaruhi harga saham, *NLP* boleh memberikan maklumat yang berharga. Tambahan pula, kerjasama dengan pakar domain boleh memberikan input yang berharga bagi memperkenalkan ciri-ciri yang lebih canggih ke dalam model ramalan. Ciri-ciri seperti tarikh pengeluaran dividen atau laporan suku tahunan syarikat boleh meningkatkan kuasa ramalan model dengan menangkap peristiwa dan faktor penting yang khusus dalam domain pasaran saham. Kemajuan ini boleh membantu meningkatkan ketepatan dan keberkesanan model ramalan harga saham.

Penghargaan

Terlebih dahulu saya ingin mengucapkan kesyukuran kepada tuhan kerana telah mengurniakan saya kesihatan dan petunjuk untuk menyiapkan penyelidikan ini dengan sempurna.

Saya ingin merakamkan ucapan jutaan terima kasih kepada penyelia projek akhir tahun saya, Prof. Dr. Shahrul Azman Mohd Noah dari Fakulti Sains dan Teknologi Maklumat (FTSM) di atas segala bimbingan dan tunjuk ajar yang diberi sepanjang tempoh pelaksanaan penyelidikan

ini. Beliau sentiasa membantu dan menyelesaikan masalah dan soalan saya dengan sabar. Selain itu, penghargaan juga ditujukan kepada Dr. Saidah Saad sebagai pemeriksa penyelidikan saya. Beliau telah memberi nasihat dan memeriksa projek tahun akhir ini sepanjang pelaksanaan penyelidikan ini.

Saya juga ingin mengucapkan jutaan terima kasih kepada ibu bapa serta ahli keluarga saya selama ini kerana banyak memberi sokongan supaya saya meneruskan pembelajaran dan melaksanakan projek tahun akhir ini. Penghargaan yang tidak terhingga juga diucapkan kepada rakan-rakan seperjuangan saya.

Akhir kata, sekali lagi saya ingin ucapkan setinggi-tinggi penghargaan dan terima kasih kepada semua pihak yang terlibat.

Sekian, terima kasih.

NG WEI SHENG

A180443

RUJUKAN

(Arielle, Chris, Sam 2022)

Arielle O'Shea, Chris Davis, Sam Taube. 2022. What Is the Stock Market and How Does It

Work? <https://www.nerdwallet.com/article/investing/what-is-the-stock->

[market#:~:text=The%20stock%20market%20is%20where%20investors%20buy%20an](https://www.nerdwallet.com/article/investing/what-is-the-stock-market#:~:text=The%20stock%20market%20is%20where%20investors%20buy%20an)

d%20sell%20shares,rather%20than%20through%20an%20exchange). [8 November 2022]

(Hongzheng Li, Shaohang Huang 2020)

Hongzheng Li, Shaohang Huang. 2020. Research on the Prediction Method of Stock Price Based on RBF Neural Network Optimization Algorithm. https://www.e3s-conferences.org/articles/e3sconf/pdf/2021/11/e3sconf_netid2021_03088.pdf

(iFAST Research Team 2021)

iFAST Research Team. 2021. The 5 key sectors in Bursa Main Market you should know. <https://www.fsmone.com.my/stocks/research/article-details/235502/the-5-key-sectors-in-bursa-main-market-you-should-know>. [8 November 2022]

(Jagruti, Sujata, Bijay, Harshvardhan 2020)

Jagruti Hota, Sujata Chakravarty, Bijay K. Paikaray dan Harshvardhan Bhojar. 2020. Stock Price Prediction Using Machine Learning Techniques <https://ceur-ws.org/Vol-3283/Paper86.pdf>.

(Mehar, Deeksha, Vinay Anand, Arun 2020)

Mehar Vijn, Deeksha Chandola, Vinay Anand Tikkiwal, Arun Kumar. 2020. Stock Closing Price Prediction Using Machine Learning Techniques. <https://www.sciencedirect.com/science/article/pii/S1877050920307924>. [16 April 2020]

(Research & Ranking 2022)

Research & Ranking. 2022. Shocking But True: 90% of People Lose Money in Stock.

<https://blog.researchandraking.com/90-percent-lose-money-in-stocks/>. [8 November 2022]

(Shreya 2022)

Shreya Pawaskar. 2022. Stock Price Predictions Using Machine Learning Algorithms.

<https://www.ijraset.com/research-paper/stock-price-prediction-using-machine-learning#conclusion>. [1 November 2022]

(Syafiqah Salim 2022)

Syafiqah Salim. 2022. Bursa Malaysia's total number of listed companies rises to 982, highest

since 2008. <https://www.theedgemarkets.com/article/bursa-malaysias-total-number-listed-companies-rises-982-highest-2008>. [8 November 2022]

(Yixin Guo 2022)

Yixin Guo. 2022. Stock Price Prediction Using Machine Learning. <https://www.diva-portal.org/smash/get/diva2:1672304/FULLTEXT01.pdf>.

Ng Wei Sheng (A180443)

Prof. Dr. Shahrul Azman Mohd Noah

Fakulti Teknologi & Sains Maklumat,

Universiti Kebangsaan Malaysia