



11th MALAYSIA STATISTICS CONFERENCE 2024

Data and Artificial Intelligence: Empowering the Future

Sasana Kijang, Bank Negara

19th September 2024

Perbandingan Model LSTM (RNN) dan Model ARIMA untuk Peramalan Harga Getah Lateks

Halimatun Sa'adiah Md Salehan¹, Noor Elidawati Bahari¹, Rahimah Md Yunos¹

¹Kolej Universiti Agrosains Malaysia, Lot 2020, Jalan Ayer Paabas, 78000, Alor Gajah, Melaka

Abstrak:

Ramalan harga getah lateks adalah penting untuk memahami dinamik pasaran dan membuat keputusan yang tepat dalam industri getah. Kajian ini membandingkan keberkesanan model *Long Short-Term Memory* (LSTM) dan *AutoRegressive Integrated Moving Average* (ARIMA) dalam meramalkan harga getah lateks menggunakan data bulanan dari 2012 hingga 2023 dengan perisian R. Model ARIMA dipilih kerana keupayaannya menganalisis data siri masa tradisional, manakala LSTM dipilih untuk kemampuannya mengendalikan data kompleks. Prestasi kedua-dua model dinilai berdasarkan *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Error* (MAE). Hasil kajian menunjukkan model LSTM lebih tepat dalam meramalkan pola jangka panjang, sementara ARIMA lebih sesuai untuk ramalan jangka pendek. Kajian ini memberi kesimpulan bahawa gabungan teknik pembelajaran mesin moden dengan kaedah statistik tradisional dapat meningkatkan ketepatan ramalan ekonomi.

Kata Kunci:

Ramalan harga; getah lateks; LSTM; ARIMA; siri masa; pembelajaran mesin; analisis statistik; perisian R.

1. Pengenalan:

Getah lateks memainkan peranan penting dalam industri global kerana penggunaannya yang meluas dalam pelbagai produk, termasuk tayar, sarung tangan, dan produk perubatan. Oleh itu, ramalan harga getah lateks yang tepat adalah kritikal untuk memastikan keputusan yang tepat dan berkesan dalam rangkaian bekalan dan pengeluaran. Keupayaan untuk meramalkan harga getah lateks dengan tepat dapat membantu pihak berkepentingan membuat keputusan yang lebih berkesan, meminimumkan risiko kewangan, dan meningkatkan kecekapan operasi [1][2].

Kajian ini bertujuan untuk membandingkan keberkesanan dua model ramalan harga getah lateks, iaitu Model *Long Short-Term Memory* (LSTM) dan Model *AutoRegressive Integrated Moving Average* (ARIMA). Model LSTM merupakan model dari jenis *Recurrent Neural Network* (RNN) yang berfungsi dalam menangani data siri masa yang kompleks,

manakala model ARIMA adalah kaedah statistik tradisional yang sering digunakan dalam analisis siri masa [3][4]. Perbandingan ini adalah penting untuk menentukan model yang lebih sesuai dalam meramalkan harga getah lateks, seterusnya membantu dalam membuat keputusan yang lebih berkesan dalam industri getah [5].

Dalam kajian terdahulu, model ARIMA telah menunjukkan keberkesanan yang tinggi dalam ramalan jangka masa pendek bagi data siri masa. Model ini bergantung kepada corak autokorelasi dalam data untuk membuat ramalan [6][7]. Walau bagaimanapun, model ini mungkin tidak mampu menangani data siri masa yang berubah dengan ketara. Kelemahan ini memerlukan pengembangan dan penerokaan model yang lebih fleksibel dan mampu menangani perubahan dinamik dalam data [8].

Sebaliknya, model LSTM telah banyak digunakan dalam kajian terkini kerana kemampuannya menangani data yang tidak stasioner dan mengenal pasti pola bagi jangka masa panjang. LSTM menggunakan memori jangka panjang untuk menyimpan maklumat penting dan membuang maklumat yang tidak berkaitan, menjadikannya lebih sesuai untuk data yang mempunyai pola yang kompleks dan tidak linear [9][10]. Keupayaan LSTM untuk belajar dan mengingat maklumat dalam jangka masa yang panjang menjadikannya model yang berpotensi untuk ramalan harga getah lateks yang lebih tepat [11].

Kajian ini menggunakan data harga getah lateks bulanan dari tahun 2012 hingga 2023. Data tersebut diproses dan dianalisis menggunakan kedua-dua model LSTM dan ARIMA. Proses ini melibatkan pengendalian nilai yang hilang, normalisasi data, serta penilaian dan perbandingan prestasi model menggunakan metrik seperti Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan Mean Absolute Error (MAE) [12]. Hasil kajian ini dijangka dapat memberikan pandangan yang lebih mendalam mengenai prestasi kedua-dua model dalam meramalkan harga getah lateks dan menyumbang kepada peningkatan ketepatan ramalan dalam aspek ekonomi.

Dengan memahami kelebihan dan kekurangan setiap model, kajian ini bertujuan untuk menyediakan asas yang kukuh bagi penggunaan model ramalan yang lebih berkesan dalam industri getah. Penemuan dari kajian ini dijangka dapat membantu industri getah dalam merancang strategi yang lebih baik dan membuat keputusan yang lebih tepat, seterusnya meningkatkan daya saing dalam pasaran global.

2. Metodologi:

2.1 Pengumpulan Data

Data harga getah lateks bulanan dari tahun 2012 hingga 2023 diperoleh dari laman web rasmi Lembaga Getah Malaysia (LGM). Data sekunder ini digunakan untuk analisis dan ramalan menggunakan model ARIMA dan LSTM. Proses ini melibatkan muat turun data dari laman web LGM serta pengesahan kesahihan dan konsistensi data untuk memastikan analisis berdasarkan maklumat yang tepat.

2.2 Pra-pemprosesan Data

Pra-pemprosesan data adalah langkah kritikal untuk memastikan kualiti dan kebolehpercayaan data yang digunakan dalam model. Langkah-langkah ini termasuk mengendalikan nilai yang hilang, normalisasi data, dan pembahagian data. Nilai yang hilang diatasi dengan kaedah interpolasi atau penghapusan rekod yang tidak lengkap [13].

Normalisasi data dilakukan untuk memastikan skala yang konsisten dan penting untuk model LSTM [14]. Normalisasi melibatkan penukaran data kepada skala antara 0 dan 1. Data kemudian dibahagikan kepada set latihan (80%) dan ujian (20%) untuk menilai prestasi model pada data yang tidak dilihat semasa latihan.

2.3 Pembinaan Model ARIMA

Model ARIMA dibina berdasarkan prinsip statistik yang kukuh. Penentuan peringkat ARIMA yang sesuai (p , d , q) dilakukan menggunakan Akaike Information Criterion (AIC) dan Bayesian Information Criterion (BIC) untuk memilih model terbaik [15]. Model dilatih menggunakan data latihan dan parameter disesuaikan untuk memaksimumkan ketepatan ramalan. Pengesahan model dilakukan dengan data ujian dan prestasi dinilai menggunakan metrik seperti *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Error* (MAE) [16].

2.4 Pembinaan Model LSTM

Model LSTM (*Long Short-Term Memory*) dibina menggunakan rangkaian neural berulang (RNN) dengan beberapa langkah penting. Reka bentuk rangkaian LSTM yang sesuai melibatkan penentuan bilangan lapisan tersembunyi dan unit LSTM. Teknik regularisasi seperti *dropout* digunakan untuk mengelakkan *overfitting* [17]. Model LSTM dilatih menggunakan data latihan, dengan parameter yang dioptimumkan melalui kaedah backpropagation. Pengesahan model dilakukan dengan menggunakan data ujian, dan prestasi model dinilai menggunakan metrik seperti MSE, RMSE, dan MAE.

2.5 Penggunaan Perisian R

Dalam kajian ini, perisian R digunakan untuk menganalisis data dan pembinaan model. R dipilih kerana keupayaannya yang tinggi dalam menangani analisis data siri masa serta ketersediaan pelbagai pakej yang sesuai untuk pembinaan model statistik seperti ARIMA dan model pembelajaran mesin seperti LSTM [18]. Dalam kajian ini, pakej-pakej R seperti '*forecast*' dan '*keras*' digunakan secara khusus untuk melatih dan mengesahkan model yang dibina.

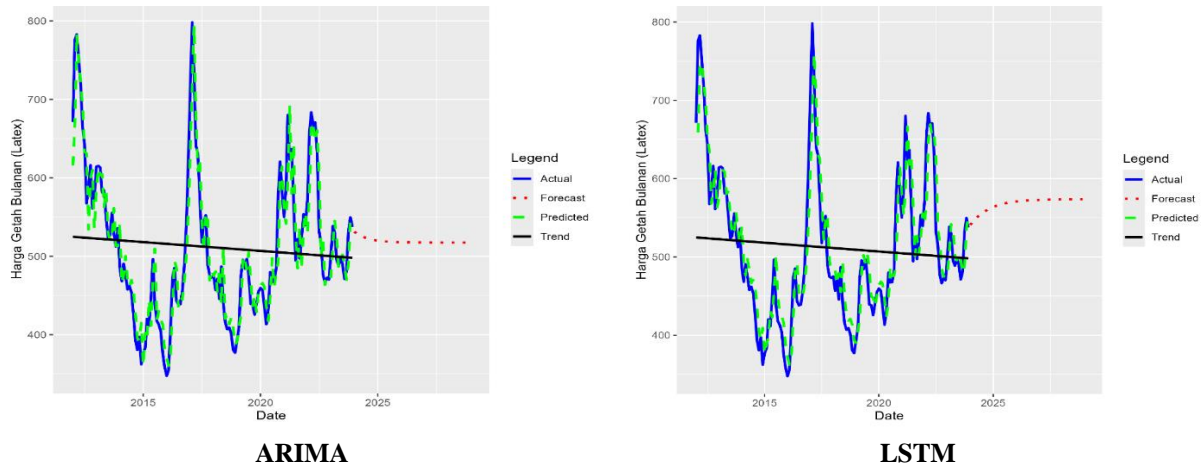
2.6 Perbandingan Prestasi Model

Kedua-dua model ARIMA dan LSTM dinilai berdasarkan prestasi mereka dalam meramalkan harga getah lateks. Perbandingan dibuat menggunakan metrik seperti *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Error* (MAE). MSE adalah purata kuasa dua perbezaan antara nilai sebenar dan nilai yang diramalkan, RMSE adalah akar kuasa dua purata perbezaan antara nilai sebenar dan nilai yang diramalkan, manakala MAE adalah purata nilai mutlak perbezaan antara nilai sebenar dan nilai yang diramalkan. Analisis perbandingan ini memberikan pandangan yang lebih mendalam mengenai prestasi setiap model dalam konteks ramalan harga getah lateks.

3. Hasil Analisis:

Model ARIMA yang terbaik dipilih dengan menggunakan *Akaike Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC) yang menghasilkan model dengan parameter ($p=2$, $d=1$, $q=2$)..Model LSTM pula telah menggunakan rangkaian neural berulang (RNN) dengan beberapa langkah penting. Reka bentuk rangkaian LSTM melibatkan dua lapisan tersembunyi dengan 50 unit LSTM setiap lapisan.

Rajah 1 menunjukkan perbandingan antara nilai sebenar dan nilai yang diramalkan oleh model ARIMA dan LSTM. Berdasarkan graf ini, dapat dilihat bahawa model ARIMA memberikan ketepatan yang lebih baik dalam jangka masa pendek dengan mengikuti corak harga sebenar dengan lebih rapat. Sebaliknya, model LSTM menunjukkan ramalan yang lebih stabil dan mencerminkan ramalan masa depan dengan lebih baik, yang menjadikannya lebih sesuai untuk ramalan jangka panjang.



Rajah 1. Perbandingan nilai sebenar dan ramalan Model ARIMA dan LSTM

Jadual 1 menunjukkan perbandingan Metrik Prestasi iaitu RMSE, MAE, R^2 , ME, MAPE dan R bagi model ARIMA dan LSTM. Berdasarkan jadual ini, model ARIMA menunjukkan prestasi yang lebih baik dalam beberapa metrik utama seperti RMSE, MAE, dan R^2 . Model ARIMA mencatat RMSE sebanyak 34.46, MAE sebanyak 26.36, dan R^2 sebanyak 0.87. Sebaliknya, model LSTM menunjukkan nilai RMSE sebanyak 36.59, MAE sebanyak 28.32, dan R^2 sebanyak 0.85. Walaupun perbezaan dalam metrik ini tidak terlalu besar, ia menunjukkan bahawa model ARIMA memberikan ketepatan yang lebih baik dalam meramalkan harga getah lateks berdasarkan data yang digunakan.

Jadual 1. Metrik Prestasi

Metrik	ARIMA	LSTM
RMSE	34.4589921	36.5883955
MAE	26.3609165	28.3167298
R^2	0.87012707	0.85170019
R	0.93285736	0.92416493
ME	-0.93271198	-4.45854316
MAPE	5.07209438	5.484089791

4. Perbincangan dan Kesimpulan:

Hasil kajian ini menunjukkan bahawa kedua-dua model ARIMA dan LSTM mempunyai kekuatan dan kelemahan masing-masing dalam meramalkan harga getah lateks. Berdasarkan Jadual 1, model ARIMA menunjukkan prestasi yang lebih baik dalam beberapa metrik utama seperti RMSE, MAE, dan R^2 . Model ARIMA mencatat RMSE sebanyak 34.46, MAE sebanyak 26.36, dan R^2 sebanyak 0.87. Sebaliknya, model LSTM menunjukkan nilai RMSE sebanyak 36.59, MAE sebanyak 28.32, dan R^2 sebanyak 0.85. Walaupun perbezaan dalam metrik ini tidak terlalu besar, ia menunjukkan bahawa model ARIMA memberikan ketepatan yang lebih baik dalam meramalkan harga getah lateks berdasarkan data yang digunakan [19].

Perbandingan metrik prestasi antara kedua-dua model menunjukkan bahawa model ARIMA mempunyai kelebihan dalam ramalan jangka pendek [20]. Model ARIMA mampu memberikan nilai ramalan yang lebih dekat dengan nilai sebenar dalam tempoh masa yang pendek. Sebaliknya, model LSTM menunjukkan keupayaan yang lebih baik dalam meramalkan pola jangka panjang. Model LSTM mampu mengenal pasti pola yang lebih kompleks dan tidak linear dalam data, yang menjadikannya lebih sesuai untuk ramalan jangka Panjang[21].

Rajah 1 menunjukkan bahawa model LSTM memberikan ramalan yang lebih stabil untuk jangka masa panjang, manakala model ARIMA menunjukkan ketepatan yang lebih baik dalam jangka masa pendek. Walaupun model LSTM tidak sebaik model ARIMA dalam beberapa metrik utama, ia masih menunjukkan prestasi yang baik dan boleh digunakan untuk ramalan jangka panjang. Model LSTM mempunyai keupayaan untuk menangani pola yang kompleks dan tidak linear dalam data, yang merupakan kelebihan besar dalam ramalan jangka panjang.

Berdasarkan hasil ini, kajian ini mencadangkan agar industri getah menggunakan model ARIMA untuk ramalan jangka pendek dan model LSTM untuk ramalan jangka Panjang [22]. Penggunaan kedua-dua model ini dapat memberikan gambaran yang lebih komprehensif dan membantu pihak berkepentingan dalam membuat keputusan yang lebih tepat[23].

Secara keseluruhannya, kajian ini menunjukkan bahawa kedua-dua model ARIMA dan LSTM mempunyai kelebihan masing-masing dan boleh digunakan untuk meningkatkan ketepatan ramalan harga getah lateks. Model ARIMA sesuai untuk ramalan jangka pendek manakala model LSTM lebih sesuai untuk ramalan jangka panjang. Dengan memahami kelebihan dan kekurangan setiap model, pihak berkepentingan dalam industri getah dapat memilih model yang paling sesuai berdasarkan keperluan ramalan mereka. Penemuan ini dijangka dapat membantu industri getah dalam merancang strategi yang lebih baik dan membuat keputusan yang lebih tepat, seterusnya meningkatkan daya saing dalam pasaran global.

Rujukan:

1. Ali, M. F., Akber, M. A., Smith, C., & Aziz, A. A. (2021). The dynamics of rubber production in Malaysia: Potential impacts, challenges and proposed interventions. *Forest Policy and Economics*, 127, 102449.
2. Ali, M. F., Aziz, A. A., & Sulong, S. H. (2020). The role of decision support systems in smallholder rubber production: Applications, limitations and future directions. *Computers and Electronics in Agriculture*, 173, 105442.
3. Bharadiya, J. P. (2023). Exploring the use of recurrent neural networks for time series forecasting. *International Journal of Innovative Science and Research Technology*, 8(5), 2023-2027.
4. Ahammad, I., Sarkar, W. A., Meem, F. A., Ferdus, J., Ahmed, M. K., Rahman, M. R., ... & Islam, M. S. (2024). Advancing Stock Market Predictions with Time Series Analysis Including LSTM and ARIMA. *Cloud Computing and Data Science*, 226-241.
5. Khin, A. A., & Thambiah, S. (2014). Forecasting analysis of price behavior: A case of Malaysian natural rubber market. *American-Eurasian Journal of Agricultural & Environmental Sciences*, 14(11), 1187-1195.

6. Elsaraiti, M., & Merabet, A. (2021). A comparative analysis of the arima and lstm predictive models and their effectiveness for predicting wind speed. *Energies*, 14(20), 6782.
7. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159-175.
8. Wang, X., Kang, Y., Hyndman, R. J., & Li, F. (2023). Distributed ARIMA models for ultra-long time series. *International Journal of Forecasting*, 39(3), 1163-1184.
9. Staudemeyer, R. C. (2015). Applying long short-term memory recurrent neural networks to intrusion detection. *South African Computer Journal*, 56(1), 136-154.
10. Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM--a tutorial into long short-term memory recurrent neural networks. *arXiv preprint arXiv:1909.09586*.
11. Phoksawat, K., Phoksawat, E., & Chanakot, B. (2023). Forecasting smoked rubber sheets price based on a deep learning model with long short-term memory. *International Journal of Electrical & Computer Engineering (2088-8708)*, 13(1).
12. Joel, L. O., Doorsamy, W., & Paul, B. S. (2022). A review of missing data handling techniques for machine learning. *International Journal of Innovative Technology and Interdisciplinary Sciences*, 5(3), 971-1005.
13. Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A survey on missing data in machine learning. *Journal of Big data*, 8, 1-37.
14. Passalis, N., Tefas, A., Kannianen, J., Gabbouj, M., & Iosifidis, A. (2019). Deep adaptive input normalization for time series forecasting. *IEEE transactions on neural networks and learning systems*, 31(9), 3760-3765.
15. Jain, G., & Mallick, B. (2017). A study of time series models ARIMA and ETS. Available at SSRN 2898968.
16. Saigal, S., & Mehrotra, D. (2012). Performance comparison of time series data using predictive data mining techniques. *Advances in Information Mining*, 4(1), 57-66.
17. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929-1958.
18. Ospina, R., Gondim, J. A., Leiva, V., & Castro, C. (2023). An overview of forecast analysis with ARIMA models during the COVID-19 pandemic: Methodology and case study in Brazil. *Mathematics*, 11(14), 3069.
19. Jong, L. J., Ismail, S., Mustapha, A., Abd Wahab, M. H., & Idrus, S. Z. S. (2020, September). The combination of autoregressive integrated moving average (arima) and support vector machines (svm) for daily rubber price forecasting. In *IOP Conference Series: Materials Science and Engineering* (Vol. 917, No. 1, p. 012044). IOP Publishing.
20. Lee, C. M., & Ko, C. N. (2011). Short-term load forecasting using lifting scheme and ARIMA models. *Expert Systems with Applications*, 38(5), 5902-5911.
21. Sagheer, A., & Kotb, M. (2019). Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing*, 323, 203-213.
22. Passalis, N., Tefas, A., Kannianen, J., Gabbouj, M., & Iosifidis, A. (2019). Deep adaptive input normalization for time series forecasting. *IEEE transactions on neural networks and learning systems*, 31(9), 3760-3765.
23. Ali, M. F., Akber, M. A., Smith, C., & Aziz, A. A. (2021). The dynamics of rubber production in Malaysia: Potential impacts, challenges and proposed interventions. *Forest Policy and Economics*, 127, 102449.

NOTE: THE REQUIRED NUMBER OF PAGES FOR PAPER IS SIX PAGES