FORECATING NATURAL RUBBER PRICE IN MALAYISA BY USING BRANCHING BI-LSTM

Ho Wei Ren[1] & Azizi Bin Abdullah[2]

[1,2]*Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia, 43600 UKM Bangi,, Selangor Darul Ehsan, Malaysia*

Abstract

Malaysia is one of the main global natural rubber producer. Rubber industry creating numerous job opportunities and contribute a huge amount of value to gross domestic product (GDP). However, there is a problem arise since the trend of SMR20 is changes frequently and volatility. Therefore ,the purpose of this project is to develop a forecasting system that can forecast the price of Standard Malaysia Rubber 20 by using deep learning algorithm. The deep leaning algorithm that used to predict price of SMR 20 in this project is Branching Bi-LSTM. Inside the branch Bi-LSTM, the coarse layer with larger memory cell will keep track of the long-term dependence, while the fine layer with smaller memory cells will focus on short term depence. The data used is obtained from the Malaysia Rubber Board and yahoo finance which is the daily price of the SMR20, the currency of USD to RM and the crude oil price in previous years. The architecture of Branching Bi-LSTM is examined to identified the most suitable branching. This model 's performance is evaluated by using regression metrics such as Mean Absolute Error(MSE) ,Root Mean Squared Error(RMSE) and Coefficient of Determination(R^2). Different deep learning models experimented, the most effective model is Branching Bi-LSTM with MSE 4.788288202367926.

Keyword: Branching, SMR 20, Bi-LSTM , natural rubber

Introduction

Natural rubber is a latex product obtained from the rubber tree (Hevea braziliensis) for industrial use. Originally in South America, natural rubber was used to make rubber balls. When Columbus explored America, he noticed the uniqueness of a ball made from the sap of a rubber tree and he brought knowledge of rubber to Europe. In 1876, Henry Nicholas Ridely, the Director of the Singapore Botanical Garden, brought seeds of "Hevea Braziliensis" to plant in Singapore and found that they grew well, and eventually these rubber seeds were brought to Kuala Kangsar and planted there, and then distributed throughout the Malay Peninsula. Hevea Braziliensis needs to be planted in hot conditions to grow well.

SMR (Standard Malaysia Rubber) is a high-quality natural rubber with technical specifications. Natural rubber is a major sector of the country as it has produced a high export value and has produced several important technologies in the sector related to rubber products. For example, Malaysia is the world's leading producer of gloves. Malaysia was the seventh-largest producer of natural rubber in the world in 2020. In 2021, Malaysia recorded a production of 469.7 thousand metric tons, down 8.7 percent compared to the previous year. (Mohd Anim Hosnam, 2019).

LSTM is a type of recurrent neural network (RNN) that can learn order dependencies in sequence prediction problems (Jason Brownlee, 2017, 2021). RNN is a neural network type that has loop that enables the network to store information. LSTM different from traditional RNNs because it has special memory cells and gate mechanisms. These components allow it to store or discard information in large datasets. The LSTM memory cell stores information to remember patterns and trends in time series.

The Branching Bi-LSTM model is used in this project to predict the price of SMR 20. Bi-LSTM unite forward and backward directions to incorporate information from the past and the future. This helps the model to improve its forecasting abilities. When training the model, there is not only training from input to output, but also training from output to input (Siami-Namini S, Tavakoli N, Namin A S, 2019). Branching Bi-LSTM refers to the type of architecture that has multiple branches in the network. Each branch has different levels of abstraction. This model was used in the study conducted by Manoj Kollama*, Dr. Ajay Joshib in 2022. Each branch can focus on capturing different aspects of the data, allowing for accurate analysis of both short-term and long-term trends.

The objectives of this study are as follows:

- To identify the number of branch to forecast the price of SMR 20
- To develop a system that forecast the price of SMR 20
- To measure the proposed algorithm

In this project, there are certain scope that need to be considered. Branching Bi-LSTM method will be used to conduct forecasts to address the issue of SMR20 price instability. Data comprising daily SMR20 prices, global crude oil prices, and RM to USD exchange rates over a 13-year period (from 2010 to 2022) will be utilized. The constraints for this project are the lack of historical data for SMR 20. The historical data available on the internet for SMR20 only starts from the year 2010. This project can help the government to make decsion and managing the risk. Furthermore, this project contribute the advancement of model architecture by employing branching .

A literature study is to identify the main research gaps based on the constructs, theories, and methods extensively used within the context of the conducted study (Paul & Criado, 2020). The study done by Nor Farah Hanim Binti Mohamad Norizan(2021) employs the ARIMA (AutoRegressive Integrated Moving Average) model to predict future monthly prices of SMR 20 natural rubber in Malaysia. The data used spans from 1995 to 2020 for predicting rubber prices over the next decade. The performance of the ARIMA model is compared with Naïve with Trend, Double Exponential Smoothing, and Holt's Winter methods, all of which are applied to a dataset obtained from the Malaysia Rubber Board. ARIMA has the smallest error measurement compared to the other techniques, indicating its suitability for this forecast. Despite ARIMA's superior performance in this study, it is not the most suitable method for long-term forecasting, as it is primarily effective for short-term predictions.

According to the study done by Manoj Kollam(2022), the author employed a novel multi-branch LSTM model for predicting earthquakes, demonstrating superior performance compared to other deep learning techniques. However, it acknowledges that the model could be improved, as it lacks spatial information which affects its mean absolute error rate.

In ther research done by Yi Chiun Fong(2022) in study the determinants of price instability in natural rubber, the result showed the most critical variables were natural rubber production, natural rubber consumption, crude oil prices, Shanghai natural rubber prices, and synthetic rubber. The study found that Natural Rubber (NR) prices are not only determined by common market forces like supply and demand but are largely driven by external factors such as global crude oil prices, real exchange rates, Synthetic Rubber (SR) prices, and the Shanghai Futures Exchange Market.

The paper by I M Md Ghani (2017) employ employ a Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model in conjunction with an Autoregressive Moving Average (ARMA) model to capture the volatility of SMR 20 prices, particularly resulting from heavy tails and volatility clustering. The authors closely examine the data series using various statistical tests, including the Phillips-Perron (PP), Augmented Dickey-Fuller (ADF), and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests. These tests indicate that the SMR 20 dataset is not consistently heteroskedastic over time. The ARMA(1,0)-GARCH(1,2) model is found to be the best for forecasting daily S.M.R 20 rubber prices for the next 20 days in the futures market. The GARCH(1,1) model highlights the importance of model selection for accurate forecasting. According to the authors, future research should focus on additional kurtosis in the determination of S.M.R 20 rubber prices, which may be caused by the impact of exceedingly high values (outliers).

The methodology used in this project is Crisp-dm. The figure 1 show the process of Crisp-dm. This model is an iterative process that allow to redirected back to previous phase when error are detected. It can start with unknown as they can iterate and replicate to gain deeper understanding of the data.

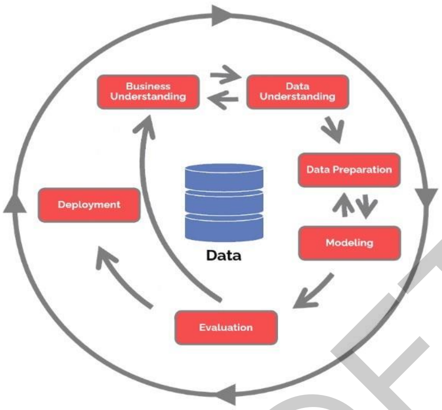


Figure 1 Process of Crisp-dm

The report is arraged in a structural manner. It start with the introduction of this project. Following by, the research methodology described the research methodology, detail of the model, dataset, data preprocessing method, analysis of data, and evaluation metrics used in the study. The results and discussion section discussed the result and suggestion for the future research.

Research Methodology

Figure 2 illustrates research methodology. At first, the dataset will undergo data preprocessing first. This step generates a data quality report, visualizes the data and applys data transformation. After the data has been preprocessed, it will divide using a 60:20:20 ratio. According to Reddy, K. S., Athelli, T., & Kulsum, S. (2022), when the size of the dataset falls within the range of 100 to 1,000,000, a division ratio of 60:20:20 should used. This ratio implies that 60% of the data is allocated to the Training Set, 20% for the validation Set, and the remaining portion to the Test Set. Following this, the model is trained using the training set. The subsequent steps involve utilizing the validation set to identify the most suitable branch architecture, most suitable optimiser and activation function. The performance of the selected model is then compared against model with other model such as LSTM sequence, Arima, CNN and other. Finally, the result by using test set is recorded.

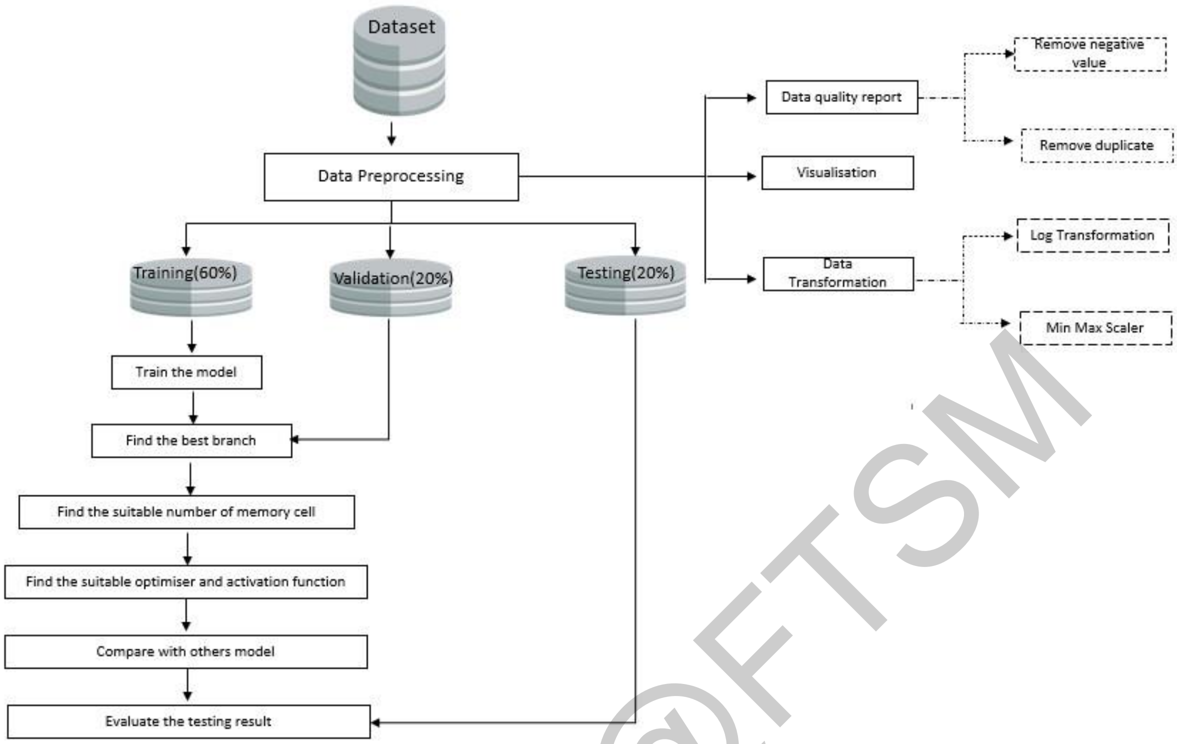


Figure 2 Research methodology model algorithm

With Bi-LSTM, the input sequence is processed by an extra LSTM layer in reverse order. Information from the past is captured by the forward LSTM layer. Future information is simultaneously captured by the backward LSTM layer. This allows the model to accurately combine context in both directions, resulting in a better understanding of sequence data. Figure 3 shows the Bi-LSTM architecture. In the figure, the LSTM layers flow in 2 directions. The forward LSTM layer focuses on historical background. The backward LSTM layer, on the other hand, captures future context. The captured information will converge in the output layer.

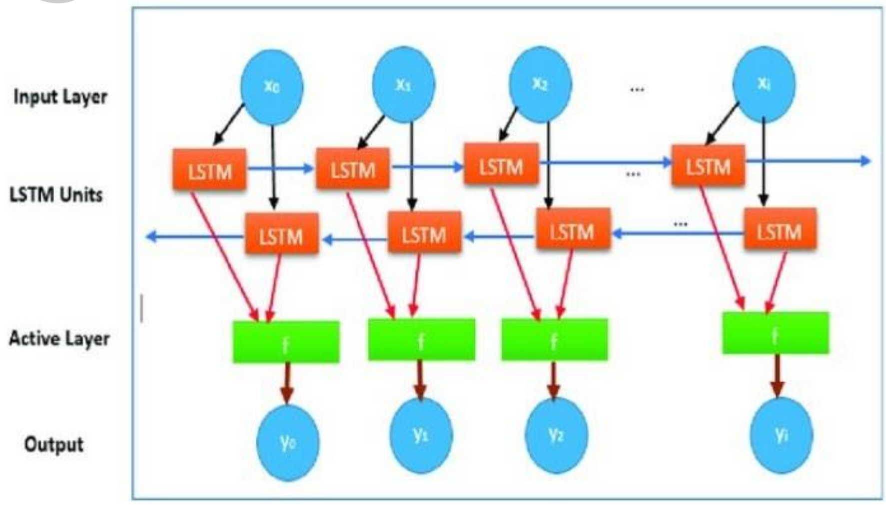


Figure 3 Architecture Bi-LSTM

Figure 4 show the architecture of branching bi-LSTM. Inputs are fed into different Bi-LSTM branches. Each branch has different layers and memory cells. This is because the coarse branch, with many memory cells, can capture the overall flow direction, while the fine branch, with fewer memory cells, can refine predictions by considering more detailed information. Prior to undergoing this process, the Bi-LSTM layers in each branch will be analyzed so the most suitable architecture can be obtained. The Branching Bi-LSTM are combined using a concatenate layer, allowing information learned at different levels of abstraction to be merged.

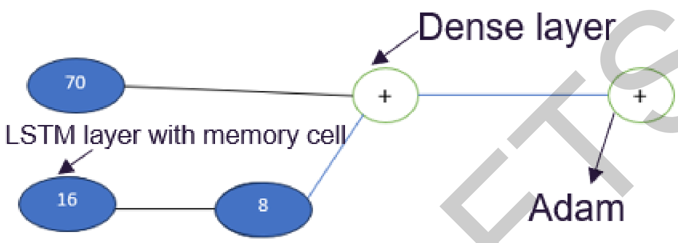


Figure 4 Architecture of branching bi-LSTM

Table 1 provides a detailed of the four features present in the dataset. The data is start from 4 january 2010 to 31 December 2022. The data was gathered from Yahoo Finance and the Malaysia Rubber Board, with the data being synchronized by date. It spans from 4th January 2010 to 31st December 2022 and includes 2,884 records.

Table 1 List of features in the dataset

Feature	Description	Data Type
Date	Corresponding date for each data entry	datetime
USD to RM currency exchange rate	The exchange rate from US dollars(USD) to Malysian Ringgit	integer
world crude oil price	Price of West Texas Intermediate (WTI) crude oil	integer
SMR 20	Price of Standard Malaysian Rubber 20 (SMR 20), a specific grade of natural rubber	integer

Table 2 summarized the methods applied in data preprocessing stage. The dataset size has been reduced from 2884 to 2776 records after the preprocessing stage.

Table 2 Data preprocessing method

Preprocessing Method	Description
Data quality report	Consist of numeric table and categorical table
Visualization	Generate visual representation
Transformation	Modify the scale of data to improve performance

The data quality report provides valuable insights for understanding the attributes within the dataset. This report includes two tables. Table 2 is for the categorical attribute, while Table 3 is for the numeric attribute. Table 2 displays the count and number of unique values for each categorical attribute. Table 3 shows the count, minimum, first percentile, median, third percentile, and maximum for each numeric attribute. From the categorical table, it is evident that the data only contains 2,877 unique entries but the total count is 2,884. This indicates that there are duplication within the data. The generated data quality report for the numeric attribute shows that the minimum value for the 'SMR 20' is negative, which is unlikely because the SMR 20 index should be non-negative.

Table 2 categorical attribute

Categorical columns:

Column	Count	Unique
date	2884	2877

Table 3 Numeric attribute

Numeric columns:

Column	Count	Mean	Std Dev	Min	25%	50%	75%	Max
usd/rm price	2884	3.7851797503467406	0.5150890303606324	2.8969	3.1997750000000003	4.0474	4.188	4.7455
wti price	2884	70.32214632454922	23.835920881109686	7.79	51.31	69.345	92.22749999999999	126.47
SMR20	2884	205.993959778086	101.49569094667049	-7.0	139.35	160.475	240.8125	569.85

Figure 5 show a histogram depicting the distribution of three variables: USD/RM exchange rate, WTI price, and SMR20. A histogram provides a visual representation of the data by breaking it down into intervals or bins and displaying either the frequency or quantity of data items in each bin. Histograms give deeper insight into the shape, spread, and central tendency of the data.

The USD/RM exchange rate exhibits a pronounced bimodal distribution. The histogram shows two peaks, one at approximately 3.2 and the other around 4.1, indicating that there are two main groups or clusters of data.

The WTI price shows a right-skewness and a distribution that is essentially normal. The histogram reveals a peak at around 50, indicating that there are many observations gathered around this value. The right skewness is due to a large number of observations scattered towards higher values.

The distribution for SMR20 also demonstrates right skewness. There is a peak in the histogram at about 150, indicating a higher concentration of observations in this area. A long tail extending towards higher values suggests that there may be a few observations with exceptionally high values. This implies that normalization needs to be carried out.

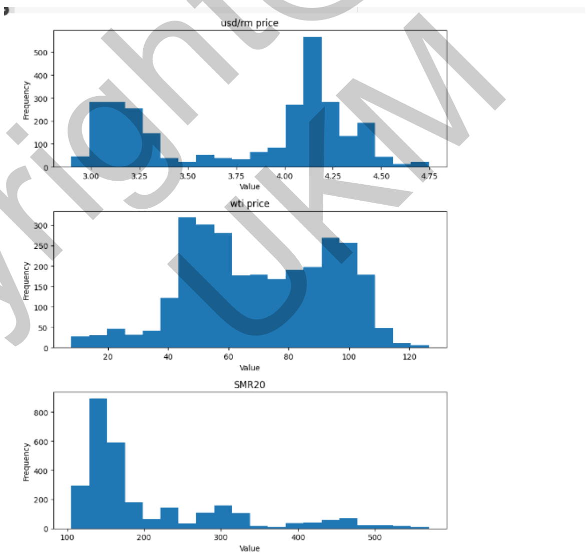


Figure 5 Histogram

Figure 6 demonstrates a Box-Whisker plot analysis, showing that the 'SMR20' column contains outliers. Outliers are data points that significantly differ from other data points in the dataset and can have a substantial impact on the results of the analysis. Outliers have been calculated using the Interquartile Range (IQR) method. In this case, only the 'SMR20' column has 244 outlier values. Goyal, C. (2021) emphasized the importance of preserving outlier values to gain a comprehensive understanding of the dataset. In a study conducted by I M Md Ghani and H A Rahim (2021), they suggested focusing on excess kurtosis in the determination of the S.M.R. 20 rubber price, which might be influenced by the effects of outlier values.

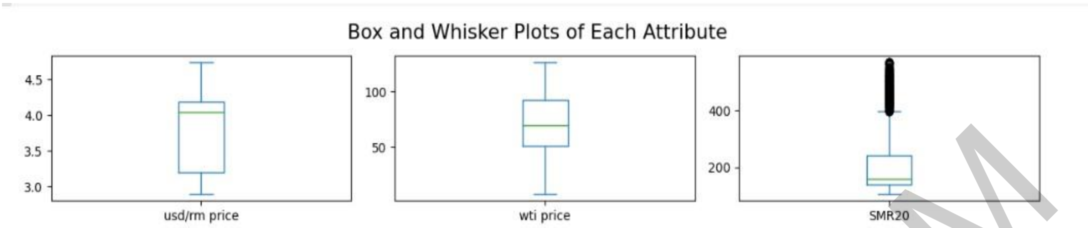


Figure 6 Box and Whisker Plots

Figure 7 displays a correlation matrix generated to examine the relationships among three variables—the USD/RM exchange rate, WTI price, and SMR20. By observing these diagrams and matrix, and relationships among the variables can be identified.

The correlation matrix between the 'USD/RM exchange rate' and 'SMR20' is -0.76, which represents a relatively strong negative correlation. This means when the 'USD/RM exchange rate' increases, the 'SMR20' value tends to decrease, and vice versa. The correlation coefficient between 'WTI price' and 'SMR20' is 0.57, representing a moderate positive correlation. This suggests that when the 'WTI price' increases, the 'SMR20' value also tends to increase, although this relationship is not as strong as the one between 'USD/RM exchange rate' and 'SMR20'.

	usd/rm price	wti price	SMR20
usd/rm price	1.000000	-0.630967	-0.760532
wti price	-0.630967	1.000000	0.572218
SMR20	-0.760532	0.572218	1.000000

Figure 7 Correlation matrix

Evaluation metrics in this project include mse, rmse and R-squared.

- Mean Squared Error(MSE)- MSE is the average squared different between the actual and predicted value
- Root Mean Squared Error(RMSE)- RMSE is the square root of mse
- R-squared(R^2)- it is the statistical measure that shows the percentage of the dependent variable's variation that can be predicted from the model's independent variables (features)

Result and Discussion

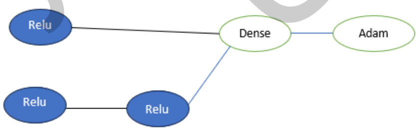
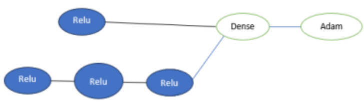
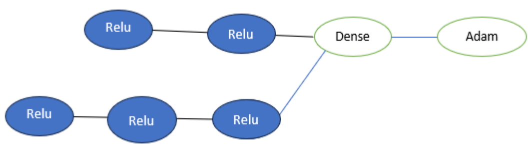
Several experiment were conducted to obtain the best forecasting model. Table 4 shows the results of techniques used to handle outliers. Outliers have a significant impact on the forecasting model as they represent influential information. The study findings indicate that the impact of outliers can be reduced more effectively through the use of logarithmic transformations compared to retaining the outliers in their original form.

Table 4 techniques used to handle outliers

Teknik	Validation result (MSE)
Retaining outliers	5.945567577181845
Logarithmic transformation	4.78288202367926

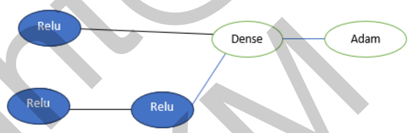
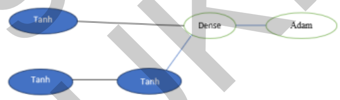
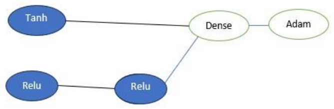
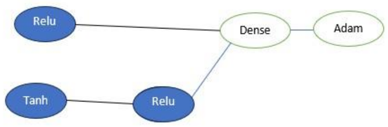
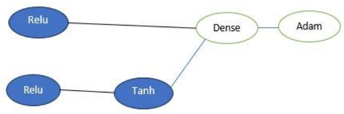
Table 5 shows the results of model architecture on performance. The results indicate that a branched architecture with 3 layers performs better than a sequential arrangement because it is more effective in capturing complex dependencies. The MSE values also decrease as the architecture becomes deeper, indicating improved performance.

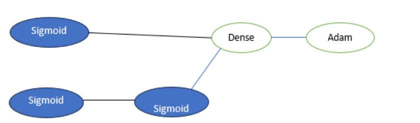
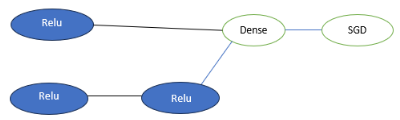
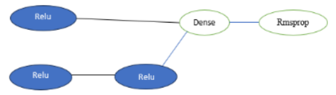
Table 5 Result of model Architecture

Model	Architecture	Hasil pengesahan (MSE)
LSTM sequence		11.449281632596007
Branching Bi-LSTM (3 layer)		4.78288202367926
Branching Bi-LSTM (4 layer)		6.064856052552339
Branching Bi-LSTM (5 layer)		5.846625891556999

The optimizer and activation function also play a crucial role in the model's performance. Table 6 shows the results of various optimizer and activation function tests. The findings indicate that the Adam optimizer provides the best performance among the considered options. Additionally, the combination of ReLU, tanh, and ReLU activation functions also shows the best results for this model. Although the combination of ReLU and tanh yields the best performance, it can lead to poor long-term predictions. The tanh function saturates at either +1 or -1, depending on whether the input is very large or very small. This causes the gradient of the function to approach zero when saturation occurs, leading to relatively small changes in the network's weights during the backpropagation process. Therefore, the combination of ReLU is chosen. This emphasizes the need to carefully select and combine activation functions to achieve better results.

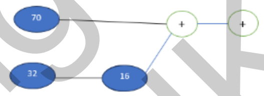
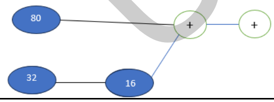
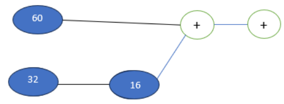
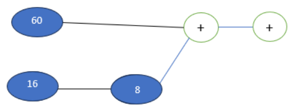
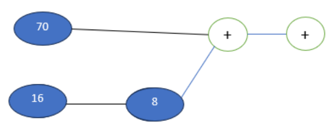
Table 6 Result of various optimizer and activation functions test

OptimiZer	Activation function	Architecture	Validation result (MSE)
Adam	Relu Relu Relu		4.78288202367926
Adam	Tanh(1) Tanh(1) Tanh(1)		4.358963536699213
Adam	Tanh(1) Relu(1) Relu(1)		3.057781933773294
Adam	Relu(1) Tanh(1) Relu (1)		5.560244399865724
Adam	RELU RELU TANH		4.450215235486973

Adam	Sigmoid Sigmoid Sigmoid		9.17595067354 6662
SGD	Relu Relu Relu		44.4753984084 9071
Rmsprop	Relu Relu Relu		6.38057090748 1785

Memory cell size influences the forecasting results. Larger memory cell is utilized for long-term predictions, whereas smaller memory cell is used to capture short-term dependencies. Table 7 presents the study results, where the optimal performance is achieved with a mixture of memory cell sizes: 70, 16, and 8.

Table 7 Result on number of memory cell

Number of memory cell	Architecture	Validation result (MSE)
70 32 16		6.064856052552339
80 32 16		5.917392657697742
60 32 16		4.8418771697273884
60 16 8		5.975970485597397
70 16 8		4.78288202367926

70 64 32		4.789423679610539

Table 8 shows a comparison of various machine learning models: CNN, LSTM(Sequential), Branch LSTM, Branch Bi-LSTM, and ARIMA, based on performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-Square, and Mean Absolute Percentage Error (MAPE). The Branch Bi-LSTM model is the best-performing model, with the lowest values for MAE, MSE, and RMSE. On the other hand, ARIMA shows the lowest performance, with a higher MSE value..

Table 8 Comparison of Various model

	MAE	MSE	RMSE	R SQUARE	MAPE
CNN	1.345435462434394	7.022965823047983	2.650087889683658	0.9041355067276741	4.076628302831708
LSTM(urutan)	2.1395015226895797	11.449281632596007	3.383678712968477	0.8898071620906435	4.844959286496973
Bi-LSTM(urutan)	1.3699161721144268	5.707716456497972	2.3890827646814525	0.9241230385308086	3.7813625066852103
Branch LSTM	1.4203443695772817	5.201717116026418	2.280727321717004	0.9367652685097104	2.8123761799294664
Branch bi-lstm	1.174737379995871	4.78288202367926	2.1869801150626085	0.951963956975621	3.129905919701128
Arima	2.4828693788843035	26.30202446214422	5.12854993756951	-	1.2040787588068225

One advanced studies have been carried out on a larger dataset to verified the efficiency of the branching technique. The larger dataset is called "Tetuan City Electricity Consumption." This dataset comprises of 10 attributes and 52,416 records, which is larger than SMR20 dataset which only has attribute and 2886 records. The results show that the Branching Bi-LSTM model has a lower MAE, MSE and RMSE compare to the CNN and LSTM in sequence. This study highlights branching technique can improve efficiency of the model, especially with larger datasets. Table 9 shows the results on the "Tetuan City Electricity Consumption".

Table 9 Advanced study on "Tetuan City Electricity Consumption"

	MAE	MSE	RMSE
LSTM sequence(3 layer)	287.85493048844535	428432.16529873555	654.5472979844433
Branch LSTM 3 layer	140.81611668575314	99849.99301928226	315.9904951407277
Branch LSTM 4 lapisan	179.77538698644537	147344.42253951283	383.8546893545953
Branch Bi-LSTM 3 layer	114.42888034168527	65958.23310575412	256.823350001035
CNN	263.6562483939357	308254.3185046789	555.2065548106208

The dashboard has been develop by using python library Dash to provide the result of forecasting SMR 20. Bootstrap and CSS are used to provide some design to the web application and make the layout of the website. There are four page in the website which are Home, Predict, Data, and Results. This website is user-friendly since all these page are linked through a navigation bar and offer clear data visualization. Besides, the users can have responsive layout change based on the screen size of the device.

Figure 8 show the interface of the Home page. This page provides essential information about SMR20 and presents the objectives of this project. Additionally, it displays relevant statistics related to the natural rubber industry.

Figure 8 the interface of the Home page.

Figure 9 shows the interface of the SMR 20 forecasting page. The graphs on this page display historical SMR 20 data from January 4, 2010, to June 30, 2023. Furthermore, it also presents the forecasted data using the Bi-LSTM and LSTM models. Additionally, this page provides the latest SMR 20 price, which is obtained using the "scrape_rubber_prices()" function that utilizes web scraping techniques to analyze the HTML content and extract relevant information.

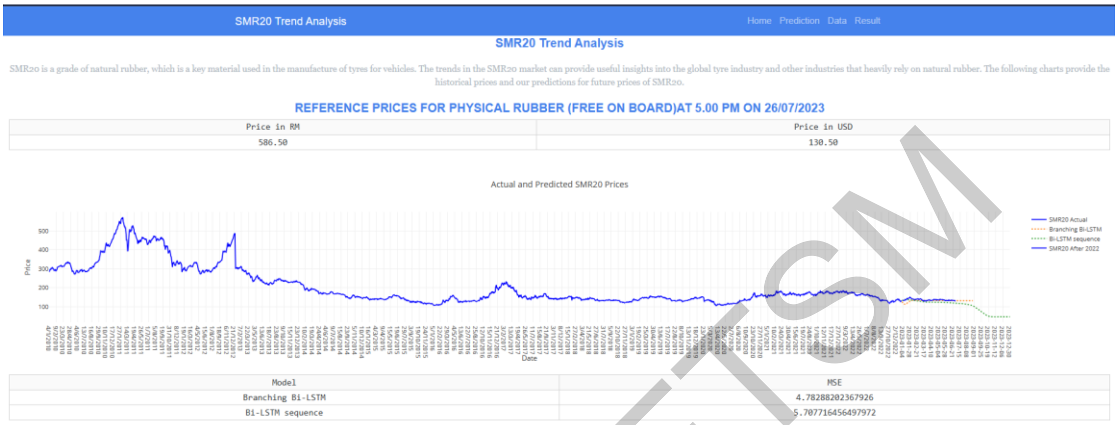
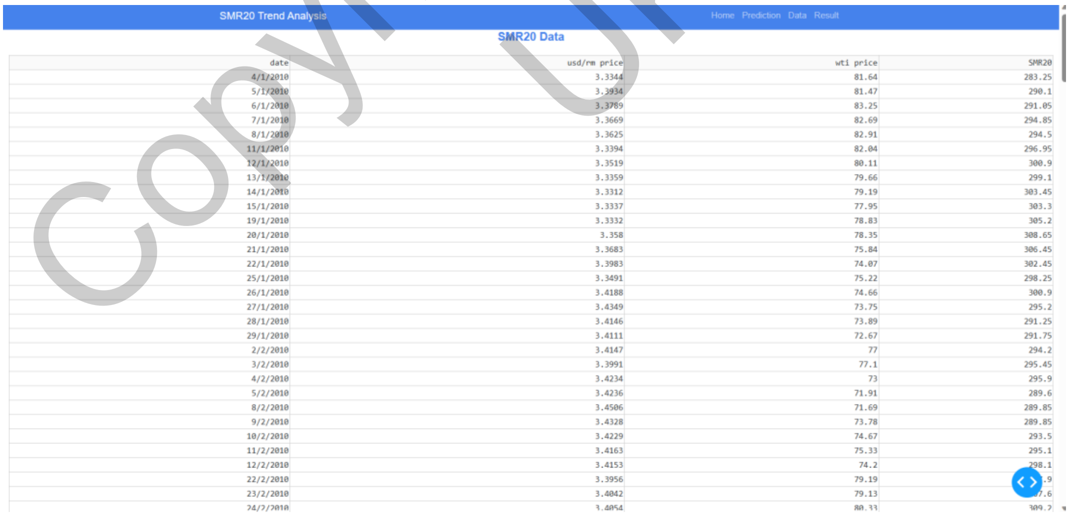


Figure 9 Interface of the SMR 20 forecasting page

Figure 10 displays the interface of the SMR 20 data page. Dash_table is utilized to create the table presented on this page. The table contains historical data for USD/MYR exchange rate, WTI price, and SMR 20 price from the year 2010 to 2022. Users are granted access to this fundamental data on this page.


SMR20 Trend Analysis

Home Prediction Data Result

SMR20 Data

date	usd/rm price	wti price	SMR20
4/1/2010	3.3344	81.64	283.25
5/1/2010	3.3934	81.47	290.1
6/1/2010	3.3789	83.25	291.05
7/1/2010	3.3669	82.69	294.85
8/1/2010	3.3625	82.91	294.5
11/1/2010	3.3394	82.04	296.95
12/1/2010	3.3519	80.11	300.9
13/1/2010	3.3359	79.66	299.1
14/1/2010	3.3312	79.19	303.45
15/1/2010	3.3337	77.95	303.3
19/1/2010	3.3332	78.83	305.2
20/1/2010	3.358	78.35	308.65
21/1/2010	3.3683	75.84	306.45
22/1/2010	3.3983	74.07	302.45
25/1/2010	3.3491	75.22	298.25
26/1/2010	3.4188	74.66	300.9
27/1/2010	3.4349	73.75	295.2
28/1/2010	3.4146	73.89	291.25
29/1/2010	3.4111	72.67	291.75
2/2/2010	3.4147	77	294.2
3/2/2010	3.3991	77.1	295.45
4/2/2010	3.4234	73	295.9
5/2/2010	3.4236	71.91	289.6
8/2/2010	3.4506	71.69	289.85
9/2/2010	3.4328	73.78	289.85
10/2/2010	3.4229	74.67	293.5
11/2/2010	3.4163	75.33	295.1
12/2/2010	3.4153	74.2	298.1
22/2/2010	3.3956	79.19	[illegible]
23/2/2010	3.4042	79.13	[illegible]
24/2/2010	3.4054	80.33	309.2

Figure 10 interface of the SMR 20 data page

Figure 11 show the interface of the results page. On this page, there is a graph displaying the differences between the actual SMR 20 prices and the forecasted prices. Additionally, a summary of the performance for different models is presented.

Model	MAE	MSE	RMSE	R SQUARE	MAPE
CNN	1.345435462434394	7.022965823047983	2.650087889683658	0.9041355067276741	4.076628302831708
LSTM (urutan)	2.1395015226895797	11.449281632596007	3.383678712968477	0.8898071620906435	4.844959286496973
Bi-LSTM (urutan)	1.3699161721144268	5.707716456497972	2.3890827646814525	0.9241230385308086	3.7813625066852103
Branch LSTM	1.4203443695772817	5.201717116026418	2.280727321717004	0.9367652685097104	2.8123761799294664
Branch bi-lstm	1.174737379995871	4.78288202367926	2.1869801150626085	0.951963956975621	3.129905919701128
Arima	2.4828693788843035	26.30202446214422	5.12854993756951	-	1.2040787588068225

Figure 11 interface of the results page

To enhance the forecasting SMR20 system,lack of historical data for SMR 20 and the limited number of attributes need to be solve. Increseasing more historical data data and attributes can produce a more robust model.

Conclusion

In conclusion, the Branching Bi-LSTM has achieve the best performance with MSE 4.78288202367926 and rmse 2.1869801150626085. The objectives have been acomplished successfully:

1. To identify the number of branch in this project

   Through heuristic experimentation, the optimal number of branch was identified. The most effective structure included two branches: with one branch with single layer and another with 2 layer. This structure show a superior performance, confirming the achieve of this objective

2. To develop a system that forecasts the price of SMR 20

   A system that can forecast SMR was successfully developed. Its predictive outcome is excellent.

3. To evaluate the proposed algorithm

   Using evaluation metrics such as MSE and RMSE the proposed algorithm demonstrate high efficiency. This confiming the objective has been successfully achieved.

This project has made successful contributions in several key areas. Firstly, it demonstrated that the branching technique can enhance performance efficiency when used in model architecture. The Branch model successfully captured various temporal relationships with different levels of abstraction.

The project also evaluated the effectiveness of conventional forecasting techniques like ARIMA and deep learning models like Branch Bi-LSTM and CNN. Clear results showed that deep learning models outperformed ARIMA in forecasting SMR 20 prices.

Lastly, but not least, the project employed log transformation techniques to address the issue of handling extremely high outlier values. The forecasting models were less affected by these high values due to the log transformation, resulting in more reliable and accurate predictions. This contribution also offers strategies that can be applied to enhance the strength and reliability of price forecasting models.

The suggestions for the further research are to increase the records in the dataset by increasing more history data because thelarger the dataset the better the model performance. Additionally, adding more attribute that reflect the economic status of America, such as interest rates. Furthermore, include attributes of economic factors related to China. This is because China is the largest consumer in the world.

Collecting more history data and include the related attribute, the model can have a better understand about the trend and can better capture the complexities. Adding attribute of economic factors from both the US and China can provide a comprehensive view of the global economic landscape. It can improve the performance of the model.

Acknowledgment

I would like to extend my heartfelt gratitude to my supervisor, Associate Professor Dr. Azizi Bin Abdullah, for his willingness to dedicate his valuable time and energy to assist and supervise me throughout this project. His expertise have been crucial in assisting me navigate through the challenges. I truly appreciate his invaluable guidance and support.

I would like to express my gratitude to the lecturers who have taught me. Thank you also to my family members who have always supported me.

RUJUKAN

Mohd Anim Hosnan. 2019. PEMISAHAN INDUSTRI KOMODITI DAN MAKANA REALITI**.** http://animhosnan.blogspot.com/2019/04/pemisahanindustri-komoditi-dan-makanan.html [27 November 2022].

Brownlee,Jason. 2017,2021 A Gentle Introduction to Long Short-Term Memory Networks by the Experts. https://machinelearningmastery.com/gentle-introductionlong-short-term-memory-networks-experts/ [27 Disember 2022].

Pirani, M. *et al.* (2022) "A comparative analysis of Arima, gru, LSTM and BILSTM on Financial Time Series forecasting," *2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)* [Preprint]. https://doi.org/10.1109/icdcece53908.2022.9793213

Yang, M. and Wang, J. (2022) "Adaptability of Financial Time Series prediction based on bilstm," *Procedia Computer Science*, 199, pp. 18–25. https://doi.org/10.1016/j.procs.2022.01.003

Paul, J. and Criado, A.R. (2020) "The Art of Writing Literature Review: What do we know and what do we need to know?," *International Business Review*, 29(4), p. 101717.https://doi.org/10.1016/j.ibusrev.2020.101717

Katariya, D. (2021) *What length of dependencies can LSTM & T-CNN really remember?*, *Medium*. Towards Data Science. Available at: https://towardsdatascience.com/how-long-dependencies-can-lstm-t-cnn-reallyremember-7095509afde8 [20 Disember2021].

Brownlee, J. (2020) *What is deep learning?,MachineLearningMastery.com*.

https://machinelearningmastery.com/what-is-deep-learning/[21 Disember 2022]

Foote, K.D. (2023) *A brief history of deep learning*, *DATAVERSITY*. Available at: https://www.dataversity.net/brief-history-deep-learning/ [22 Disember 2022].

Oppermann, A. (2022) .*Is Deep Learning and How Does It Work?* https://builtin.com/machine-learning/what-is-deep-learning (25 Disember 2022).

Mohamad Norizan, N.F. and Md Yusof, Z.B. (2021) “Forecasting natural rubber price in Malaysia by 2030” *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, 6(9), pp. 382–390. https://doi.org/10.47405/mjssh.v6i9.986

Siami-Namini, S., Tavakoli, N. and Namin, A.S. (2019) “The performance of LSTM and BiLSTM in forecasting time series,” *2019 IEEE International Conference on Big Data (Big Data)* [Preprint]. https://doi.org/10.1109/bigdata47090.2019.9005997

Wang, H. *et al.* (2023) “Dafa-BiLSTM: Deep Autoregression Feature Augmented Bidirectional LSTM network for time series prediction,” *Neural Networks*, 157, pp. 240–256. \ https://doi.org/10.1016/j.neunet.2022.10.009

Moghar, A. and Hamiche, M. (2020) “Stock market prediction using LSTM recurrent neural network,” *Procedia Computer Science*, 170, pp. 1168–1173. https://doi.org/10.1016/j.procs.2020.03.049

Foote, K.D. (2023) *A brief history of deep learning*, *DATAVERSITY*. Available at: https://www.dataversity.net/brief-history-deep-learning/ [22 Disember 2022].

Basri, M. a. M., Hapka, M. S., Jaafar, M. S., & Muhamat, A. A. (2018). Determinants of the Price of Natural Rubber in Malaysia. *The International Journal of Business and Management*, *6*(12). https://doi.org/10.24940/theijbm/2018/v6/i12/bm1812-022

KOLLAM, M. A. N. O. J., & Joshi, A. (2022). Multi-branch LSTM encoded features for forecasting earthquakes. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4124532

X. Shao, D. Ma, Y. Liu and Q. Yin, "Short-term forecast of stock price of multi-branch LSTM based on K-means," 2017 4th International Conference on Systems and Informatics (ICSAI), Hangzhou, China, 2017, pp. 1546-1551, doi: 10.1109/ICSAI.2017.8248530.

Yi Chiun Fong, Aye Aye Khin, & Chee Seong Lim. (2020). *Determinants of Natural Rubber Price Instability for Four Major Producing Countries*. Pertanika J. Soc. Sci. & Hum. 28 (2): 1179 – 1197.

Ben Ameur, H., Boubaker, S., Ftiti, Z. *et al.* Forecasting commodity prices: empirical evidence using deep learning tools. *Ann Oper Res* (2023). https://doi.org/10.1007/s10479-022-05076-6

I M Md Ghani and H A Rahim. "Modeling and Forecasting of Volatility using ARMA-GARCH: Case Study on Malaysia Natural Rubber Prices." IOP Conference Series: Materials Science and Engineering, vol. 548, 012023, 2019. doi:10.1088/1757-899X/548/1/012023. [15 Jun 2023].

Goyal, C. (2021, May 16). Why You Shouldn't Just Delete Outliers. Analytics Vidhya. Retrieved from https://www.analyticsvidhya.com/blog/2021/05/why-you-shouldnt-just-delete-outliers[3 Julai 2023].

Reddy, K. S., Athelli, T., & Kulsum, S. (2022). Analysis and Prediction of House Price. Journal of Emerging Technologies and Innovative Research (JETIR), 9(5), Retrieved from www.jetir.org (ISSN-2349-5162). [10 Julai 2023]

Ho Wei Ren (A183805)
Prof. Madya Dr. Azizi Bin Abdullah
Fakulti Teknologi & Sains Maklumat,
Universiti Kebangsaan Malaysia