

Netflix Stock Price Trend Prediction Using Recurrent Neural Network

Irani Hoeronis

Informatics, Engineering Faculty, Siliwangi University, Indonesia
iranihoeronis@unsil.ac.id

Abstract— Stocks are investments that have dynamic movements. Stock price changes move every day even hourly. With very fast changes, stock prices require predictions to be able to determine stock market projections. Predictions are used to reduce risk when making transactions. In this study, predictions of stock price trends were made using the Recurrent Neural Network (RNN). The approach taken is to perform a time series analysis using the RNN variance, namely Long Short Term Memory (LSTM). Hyperparameter construction in the LSTM model testing simulation can estimate stock prices with maximum percentage accuracy. The results showed that the prediction model produced a loss function of 0.0012 and a training time of 73 m/step. The evaluation was carried out with the RMSE which resulted in a score of 17.13325. Predictions are obtained after doing machine learning using 1239 data. The RMSE and LSTM models are calculated by changing the number of epochs, the variation between the predicted stock price and the current stock price. Computations are carried out using a stock market dataset that includes open, high, low, close, adj prices, closes, and volumes. The main objective of this study is to determine the extent to which the LSTM algorithm anticipates stock market prices with better accuracy. Code can be seen at iranihoeronis/RNN-LSTM (github.com)

Keywords— Stock Prediction, Time Series, Recurrent Neural Network (RNN), Long Short Term Memory (LSTM).

Received March 2022 / **Revised** December 2022 / **Accepted** December 2022

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



I. INTRODUCTION

Stock is an investment that provides an opportunity to get a high return, short, and fast [1]. To make investment decisions, it is necessary to know the performance and prospects of the company [2]. One of the media to see the development of stock prices is the Composite Stock Price Index (IHSG) which can reflect the enthusiasm of the market towards rising stock prices and vice versa [3].

Changes in stock prices are not easy to predict because they consider several factors that influence them, including technical factors, fundamentals, and sentiment [4]. The technique to see the movement of stock prices by observing prices in the past is a technical factor. The stock price movement is used as the basis for making predictions.

The methods and techniques used to predict stock prices are obtained from data on open, high, low, close, adj prices. close, and the volume changes every day. Open is the opening price. High is the highest price. Low is the lowest price. Close is the closing price. adj. close is the closing price that has been adjusted for dividends and stock splits. Volume is the number of exchanges traded.

Prediction of stock prices can be done using technology in the form of machine learning. The data obtained by the machine is then processed by determining the quality and quantity of learning.

The artificial neural network approach is widely used and has been used as a Machine Learning implementation technique [5]. The use of artificial neural networks represents complex input-output connectivity. In addition, ANN also can solve problems through data input and speed in execution, as well as in initiating complex systems [6].

A Recurrent Neural Network as part of an artificial neural network uses repetitive processing techniques to process input in the form of sequential data. The use of RNN as a method for classifying and identifying has been widely carried out [7][8][9]. In addition, RNN is also used to predict [10][11][12][13]. Prediction can be done using two approaches, namely the time-series approach or the cause-effects approach. The time-series approach is a model that looks at the trend of available data from the past, while the cause-effects approach is an approach that explains the causal relationship of the occurrence of a situation.

DOI : 10.24014/coreit.v8i2.16599

The stock prediction was carried out by [10] by analyzing by looking at the history of stock prices using RNN which resulted in training data accuracy of 94% and test data of 55%. In addition [14] conducted a multivariate time series prediction study using RNN for meteorological data which resulted in R2 with a value of 90.9% on monthly data of silver station 1, 89.6% on monthly data of silver station 2, 92.1% on daily data silver station 1, and 94.1% on daily data silver station 2.

Research [7] conducted a classification of quail egg quality using the RNN method with prediction precision above 75%, recall above 81%, and accuracy above 87%. Prediction of multi-time steps with RNN LSTM [13] gets RMSE results of 6888.37 on training data and 14684.33 on testing data.

The results of previous studies show that RNN is very good for making predictions based on specified variables. This study predicts stock prices with a time series approach using RNN. This paper contributes to the prediction of Netflix stocks for the last 5 years using the time series approach and evaluates the RNN model with the RMSE approach.

This paper consists of 5 chapters, namely chapter 1 contains an introduction explaining the background of the research, chapter 2 contains the research methodology used, chapter 3 describes the results and discussion and chapter 4 contains conclusions and suggestions.

II. RESEARCH METHODOLOGY

The stages of research were carried out in 3 stages, namely: preprocessing, building the RNN, and prediction and visualization. The research method starts from the first stage, namely preprocessing data then importing training sets, feature scaling, creating data structures, and reshaping. The second stage is building the RNN by importing the hardware libraries and packages, initializing the RNN, adding the first LSTM layer and some dropout regularization, adding the second, third, fourth, and output layers, compiling the RNN, fitting the RNN into the training set. The third stage is making predictions and visualization of results, namely by getting the Netflix stock price, getting the 2017-2022 Netflix stock price prediction, visualizing the results, and evaluating the RMSE. The research method for Netflix stock price prediction is depicted in Figure 1.

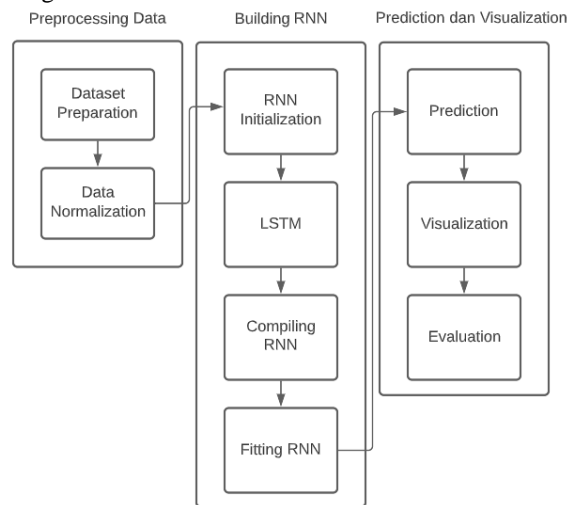


Figure 1. Netflix Stock Price Prediction Research Method

Data preprocessing is done by preparing datasets and data normalization. The dataset is divided into two parts, namely training data and test data. The preprocessing stage is built from a stock price dataset in the form of time series data sequences. Time series data is a series of data in the form of a time sequence according to the characteristics of the data taken from a set of events at a certain period of time. The dataset attributes in this study are described in table 1.

Tabel 1. Atribut Dataset

No.	Atribut	Nilai Atribut
1	Date	2017-01-04; 2018-02-14; dll
2	Open	130.1699981689453; 269.8800048828125; dll
3	High	126.55000305175781; 260.3299865722656; dll

2.1. Preprocessing data

The data preprocessing stage is the initial stage in this research. Data preparation is done by preparing several datasets, namely training data and test data, and then stored in the form of CSV (comma-separated values) files. Furthermore, in the preprocessing process, feature scaling is performed to normalize the data. The range of values (scale) used in [0,1] or [-1,1]. This is done to get faster convergence. In each value range, the feature transformation process is based on formula (1).

$$xnorm = x - \frac{\min(x)}{\max(x) - \min(x)} \quad (1)$$

Description:

$xnorm$: normalization value result

x : real/actual value to be normalized

$\max(x)$: maximum value of real/actual data

$\min(x)$: minimum value of real/actual data

In this study, the normalization process was carried out using MinMaxScaler in the sklearn library. This is done by changing the dataset to a scale of 0 (min) to 1 (max) and then performing a fit transform on the training set. After transforming the train scale, then create a train data structure so that the training process is faster. The final stage of preprocessing is reshaping the data, namely changing the rows and columns in the train dataset.

2.2. Building RNN

The method applied is to use the Recurrent Neural Network. In RNN, each input will go through the hidden layer and output the output, then it will be repeated as input again for processing in the hidden layer until the desired output is obtained.

LSTM is an RNN variance that has more accurate information capabilities [15]. One of the capabilities of the LSTM is to store errors that are generated when doing backpropagation so that errors may not increase.

Long Short Term Memory (LSTM) network aims to overcome hidden layer problems. The LSTM design incorporates non-linear, dependent data into the RNN [16]. In addition, LSTM is also a solution when processing sequential data in overcoming the vanishing gradient in RNN.

The failure of RNN to capture long term dependencies is caused by problems with vanishing gradients [17]. This can result in reduced accuracy of an RNN prediction [18]. The LSTM architecture consisting of input, output, and hidden layers is described in Figure 2.

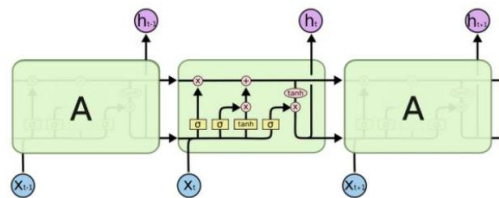


Figure 2. LSTM Architecture

Figure 2 is an LSTM architecture that presents the contents of the hidden LSTM layer, namely a memory cell that stores a state or value (cell state). The gate in an LSTM memory cell consists of an input gate, a forget gate, and an output. The input gate (i_t) is obtained from the previous output along with the new input and then passed to the sigmoid layer. Gate will return a value of 1 or 0.

$$i_t = \sigma(W_i S_{t-1} + W_i X_t) \quad (2)$$

W_i is the weight of the input gate. S_{t-1} is the state at time $t-1$. X_t is the input at time t . (σ is the sigmoid activation function).

Forget gate (f_t) is the application of the sigmoid activation function that combines the sigmoid layer of the output at time $t-1$ and input at time t .

$$f_t = \sigma(W_f S_{t-1} + W_f X_t) \quad (3)$$

W_f is the weight of the forget gate. S_{t-1} represents the state at time $t-1$. X_t is the input at time t . (σ is the sigmoid activation function).

The output gate (o_t) controls the number of states that pass through the output and does the same with other gates by generating a new cell state (h_t).

$$o_t = \sigma(W_o S_{t-1} + W_o X_t) \quad (4)$$

$$h_t = o_t * \tanh(c_t) \quad (5)$$

W_o is the weight of the output gate. S_{t-1} represents the state at time $t-1$. X_t is the input at time t . (σ is the sigmoid activation function).

2.3. Prediction and Visualization

After training on Netflix stock price data, predictions are made using test data. The determination of prediction accuracy is based on the amount of loss generated. The smaller the loss, the better. This shows how accurate the process is. The 2017-2022 Netflix stock price is depicted in figure 3.

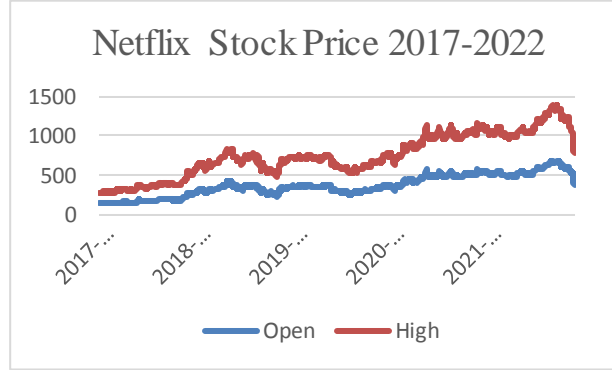


Figure 3. The 2017-2022 Netflix Stock Price

After that, visualization is done to see the actual conditions with Netflix stock price predictions in graphic form. The condition of the Netflix stock price against time is depicted in different lines and colors.

To measure the success rate of the model, an evaluation of the model is carried out. The Root Mean Squared Error (RMSE) method is used to measure the accuracy of the prediction results by calculating the percentage error between the real/actual value and the predicted value. RMSE is calculated using the following formula.

$$RMSE = \sqrt{\sum \frac{(y_{pred} - y_{ref})^2}{N}} \quad (6)$$

y_{pred} is the predicted value. y_{ref} is the real value, and N is the number of data.

III. RESULTS AND DISCUSSION

3.1. Dataset Preparation

The dataset used is Netflix stock price data for 5 years, starting from 2017 to 2022. The number of datasets consists of 1259 data which are divided into train data and test data.

The data structure is divided into 60 timesteps with 1 result output. Then reshaping the data is also carried out to modify the dimensions in rows and columns. The data structure is built with 60 timesteps and 1 output. The data is divided into x_{train} and y_{train} in the range $i=60$ to 1239. x_{train} and y_{train} are appended to the x_{train} row scale in the range $i-60:i$ and the first column. While the y_{train} training set scale is as much as the first row and column. Reshaping the x_{train} and y_{train} arrays was carried out to modify the array dimensions to (1179,60,1). The results of normalization of open and high price data are (1239,1) with a range of values explained in table 2.

Table 2. Normalized Value

Scaled Feature Value
[[0.00861288]
[0.00396266]
[0.00620634]
...
[0.68407342]
[0.47179098]

[0.4418992]]

3.2. Implementation of Data Training

The dataset that has been divided into train and test is tested by building the LSTM layer. The number of layers to test the data is 4 layers. The LSTM neuron unit is set at 50 for each layer and the dropout is 0.2. The summary of the last 5 training data for train data of 100 epochs and batch size of 32 is presented in table 3. From the training data trial, a loss of 0.0012 was obtained with a training time of 73 m/step.

Table 3. Summary of Training Results

Epoch	Time	Loss
95/100	71ms/step	0.0015
96/100	70ms/step	0.0012
97/100	72ms/step	0.0014
98/100	75ms/step	0.0012
99/100	73ms/step	0.0012
100/100	73ms/step	0.0012

3.3. Prediction Results

Netflix stock price predictions compare to actual prices. Visualization in graphic form shows that the actual condition of the Netflix stock price is not much different from the predicted condition of the Netflix stock price. The Netflix stock price prediction graph is depicted in Figure 3.

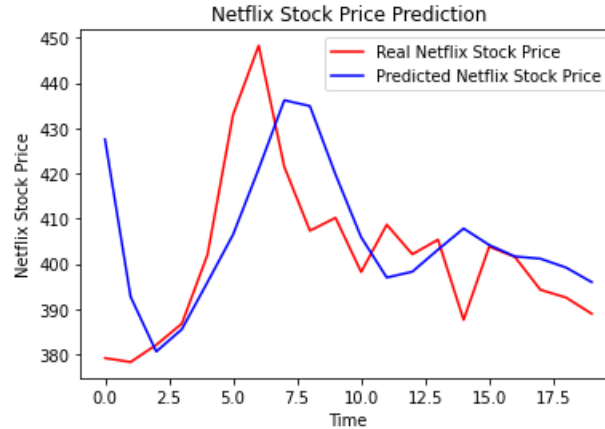


Figure 3. Netflix Stock Price Prediction

The graph shows the actual data against predictions with fit conditions. Based on the graph, it can be concluded that the model made can predict Netflix stock prices well.

3.4. Model Performance Evaluation

The RNN model built is based on a regressor to predict the sustainability of Netflix's stock price outcome. The way to evaluate the performance of the model is to use a matrix called RMSE (Root Mean Squared Error). The calculation is done by calculating the predicted value with the actual value. The results of the RMSE calculation obtained a value of 17.13325.

IV. CONCLUSION

4.1. Conclusion

This study predicts the time series of netflix stock prices for 5 years starting from 2017 to 2022. The conclusions of the research results using the RNN and LSTM model are as follows.

- The use of RNN and LSTM in making predictions results in a loss function of 0.0012 and a training time of 73 m/step
- Evaluation using RMSE resulted in a value of 17.13325

4.2. Suggestion

Addition of datasets can be done to see data trends and increase prediction accuracy. Increasing the timestep so that it can take longer is also possible. In this study the timestep used was 60 days (3 months),

it could be tried to be 120 days (6 months). In addition, more LSTM layers and neurons can be added. In this study, 4 layers and 50 LSTM neurons were used.

REFERENCES

- [1] Dini. Siti, dkk., Faktor-faktor yang Memengaruhi Harga Saham, *Jurnal Ilmiah MEA (Manajemen, Ekonomi, dan Akuntansi)*. Vol.4 No.1, STIE Muhammadiyah Bandung, 2020.
- [2] Dewi. Irma Nurmala, Pengaruh Tingkat Profitabilitas terhadap Harga Saham dan Implikasinya pada Indeks Saham di Bursa Efek Indonesia, *Jurnal Ilmu Manajemen dan Bisnis*, Vol.02 No.3, UPI Bandung, 2001.
- [3] E. Mala, S. Rochman, and A. Djunaidy, Prediksi Harga Saham yang Mempertimbangkan Faktor Eksternal Menggunakan Jaringan Saraf Tiruan, Vol.1 No.2. *Jurnal Ilmiah Nero Sidoarjo*, 2014.
- [4] L. Troiano, E.M. Villa, and V. Loia, Replicating a Trading Strategy by Means of LSTM for Financial Industry Applications, *IEEE Transactions on Industrial Informatics*, Vol.14 No.7, 2018.
- [5] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, dan G.Z. Yang, Deep Learning for Health Informatics, *IEEE Journal of Biomedical and Health Informatics*, Vol. 21 No.1 Jan 2017
- [6] Norhamreeza, Abdul Hamid, N. M. N., Rozaida Ghazali, Mohd. Najib Mohd Salleh, Accelerating Learning Performance of Back Propagation Algorithm by Using Adaptive Gain Together with Adaptive Momentum and Adaptive Learning Rate on Classification Problems, *International Journal of Software Engineering and Its Applications* Vol.5 No.4. 2011.
- [7] Tarkus, Exel Defrisco, Sompie, Sherwin R.U.A., Jacobus, Agustinus, Implementasi Metode Recurrent Neural Network pada Pengklasifikasian Kualitas Telur Puyuh, *Jurnal Teknik Informatika Unsrat* Vo.15 No.2, 2020.
- [8] Perdana, Kevin, Zulfachmi, Huda, Dwi Nurul, Identifikasi Berita Hoax dengan Rexurrent Neural Network, *Jurnal Bangkit Indonesia Tanjungpinang*, Vol.10 No.2, 2020.
- [9] Firmansyah, Muhammad Rizal, Ilyas, Ridwan, Kasyidi, Fatan, Klasifikasi Kalimat Ilmiah Menggunakan Recurrent Neural Network, *Prosiding The 11th Industrial Research Workshop and National Seminar Bandung*, 2020.
- [10] Suyudi, M. Abdul Dwiyanto Suyudi, Djamal, Esmeralda C., Maspupah, Asri, Prediksi Harga Saham Menggunakan Metode Recurrent Neural Network, *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*, 2019.
- [11] Diponegoro, Muhammad Haris, Kusumawardani, Sri Suning, Hidayah, Indriana, Tinjauan Pustaka Sistematis: Implementasi Metode Deep Learning pada Prediksi Kinerja Murid, *Jurnal Nasional Teknik Elektro dan Teknologi Informasi* Vol.10 No.2, UGM, 2021.
- [12] Rizal, Ahmad Ashril, Hartati, Sri, Prediksi Kunjungan Wisatawan dengan Recurrent Neural Network Extended Kalman Filter, *Jurnal Ilmiah Ilmu Komputer* Vol. X No.1, Universitas Udayana, 2017.
- [13] Rizal, Ahmad Ashril, Soraya, Siti, Multi Time Steps Prediction dengan Recurrent Neural Network Long Short Term Memory, *Jurnal Matrik* Vol. 18 No.1, Universitas Bumi Gora, 2018.
- [14] Hariadi, Victor, dkk., Multivariate Time Series Forecasting Using Recurrent Neural Networks for Meteorological Data, *Prosiding Seminar Nasional Teknologi Informasi dan Kedirgantaraan, Senatik* Vol V Yogyakarta, 2019.
- [15] Zhang, Lin, *Theory, Methodology, Tools, and Applications for Modelling and Simulation of Complex Systems*, Springer Business Media, Singapore, 2016.
- [16] L. Gao, Z. Guo, H. Zhang, X. Xu, and H. T. Shen, Video Captioning with Attention-Based LSTM and Semantic Consistency, *IEEE Transactions on Multimedia*, Vol.19 No.9, 2017.
- [17] Chung, H., Shin, K.S., Genetic algorithm-optimized long short-term memory network for stock market prediction, *Sustainability*, 10(10), 2018.
- [18] Zhao, Z., Chen, W., Wu, X., Chen, P.C., Liu, J., LSTM Network: A Deep Learning Approach for Short-Term Traffic Forecast, *IET Intelligent Transport Systems*, 11(2), 2017.