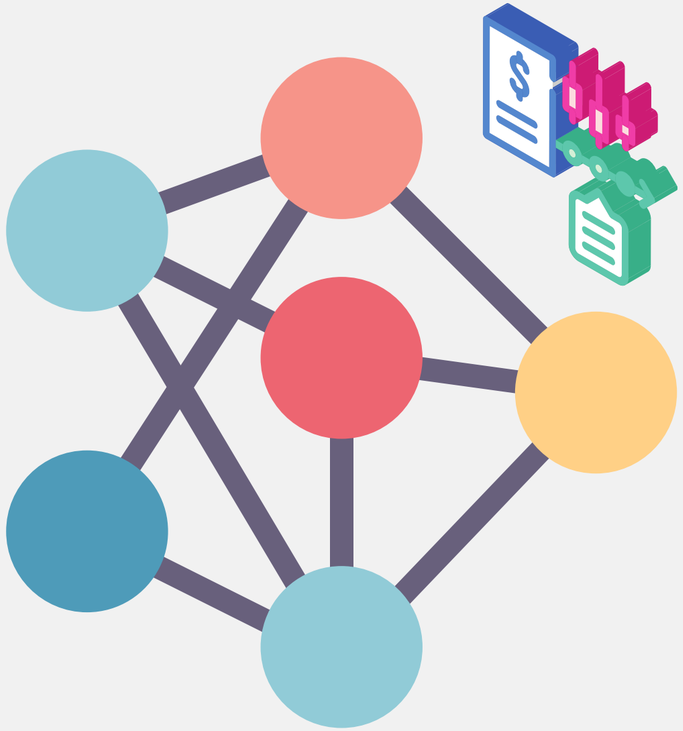
SIT723 - 2022 RESEARCH PROJECT

Financial Time Series Analysis with Machine Learning



Supervisor: Prof Maia Angelova

Research Paper

Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models

Sidra Mehtab, Jaydip Sen and Abhishek Dutta

The paper consists of **eight regression models** using the training data that consisted of NIFTY 50 index records during December 29, 2014 till December 28, 2018.

Using these regression models, the paper consists of prediction of open values of NIFTY 50 for the period December 31, 2018 till July 31, 2020

Result: The results clearly indicate that the LSTM-based univariate model that uses one-week prior data as input for predicting the next week's open value of the NIFTY 50 time series is the most accurate model.

Variable Defined

1) high_norm is computed as: high_norm = (high - Hmin)/(Hmax – Hmin).

2) low_norm = (low – Lmin)/(Lmax - Lmin). The values of low_norm also lie in the interval [0, 1].

3) close_norm = (close - Cmin) / (Cmax – Cmin)

4) volume_norm and range_norm

Case 1 - December 29, 2014 till December 28, 2018

Evaluate the training performance of the machine learning-based regression models. The predictions are made on daily basis

Case 2 - December 31, 2018 till July 31, 2020

Evaluate in terms of their prediction accuracy of open values for each of the 415 days included in the test dataset

Machine learning models

(i) multivariate linear regression
(ii) multivariate adaptive regression spline (MARS)
(iii) regression tree
(iv) bootstrap aggregation (Bagging)
(v) extreme gradient boosting (XGBoost)
(vi) random forest (RF)
(vii) artificial neural network (ANN)
(viii) support vector machine (SVM).

Result

The multivariate regression and the random forest regression were the most accurate models in terms of their forecasting accuracies on the NIFTY 50 time series.

Table 1. Multivariate regression results

Stock	Case I Training Data		Case II Test Data	
NIFTY 50	Correlation	0.99	Correlation	0.99
	RMSE	0.27	RMSE	0.42

Table 2. MARS regression results

Stock	Case I Training Data		Case II Test Data	
NIFTY 50	Correlation	0.99	Correlation	0.99
	RMSE	0.42	RMSE	0.85

Table 3. Decision tree regression results

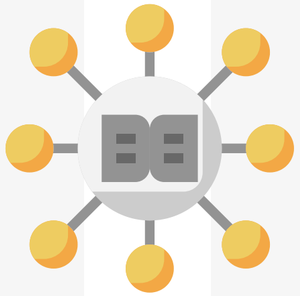
Stock	Case I Training Data		Case II Test Data	
NIFTY 50	Correlation	0.98	Correlation	0.16
	RMSE	2.52	RMSE	10.40

Table 4. Bagging regression results

Stock	Case I Training Data		Case II Test Data	
NIFTY 50	Correlation	0.99	Correlation	0.96
	RMSE	1.75	RMSE	3.72

Deep learning Models

(i) LSTM model for multi-step forecasting with univariate input data of one week,
(ii) LSTM model for multi-step forecasting with univariate input data of two weeks,
(iii) Encoder-decoder LSTM for multi-step forecasting with univariate input data for two weeks, and
(iv) Encoder-decoder LSTM for multi-step forecasting with multivariate input data for two weeks

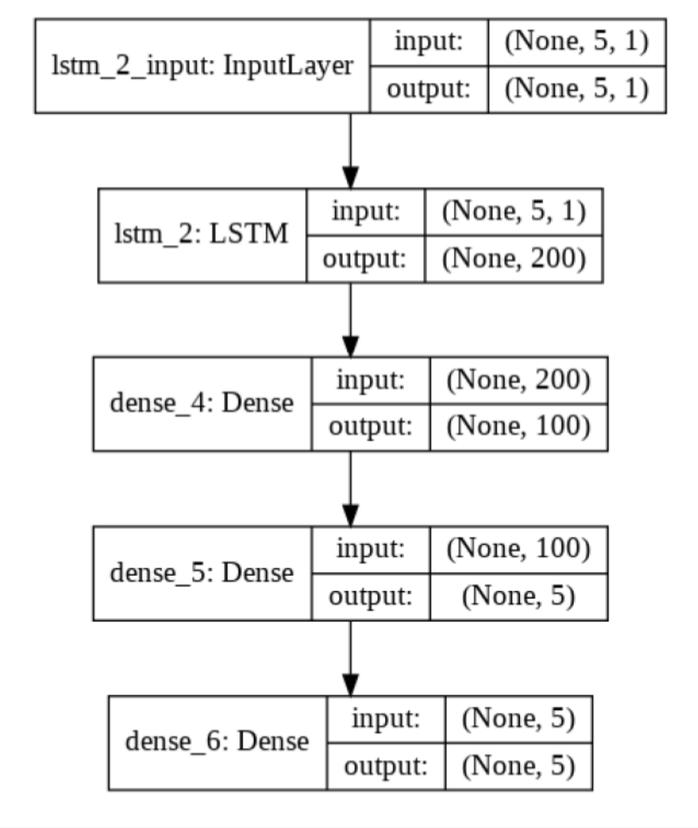


Result

The univariate LSTM model with one-week data as the input turned out to be the most optimum model – both in terms of accuracy and execution time.

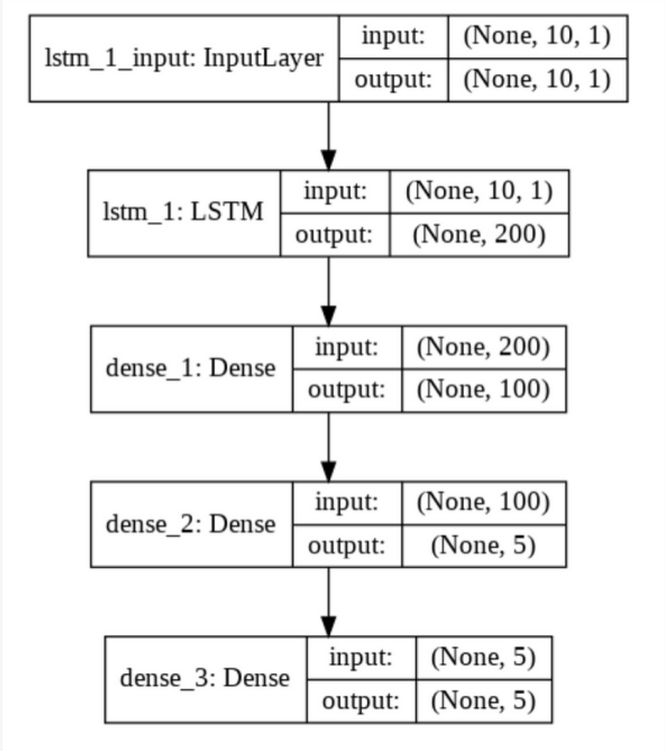
LSTM model for multi-step forecasting with univariate input data of one week

①



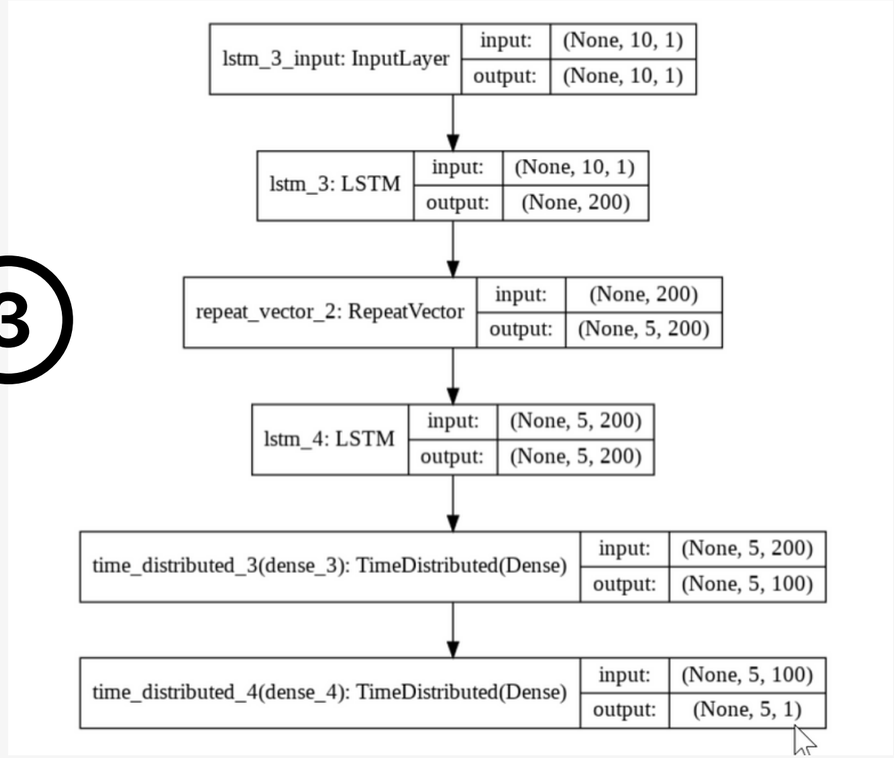
LSTM model for multi-step forecasting with univariate input data of two weeks

②



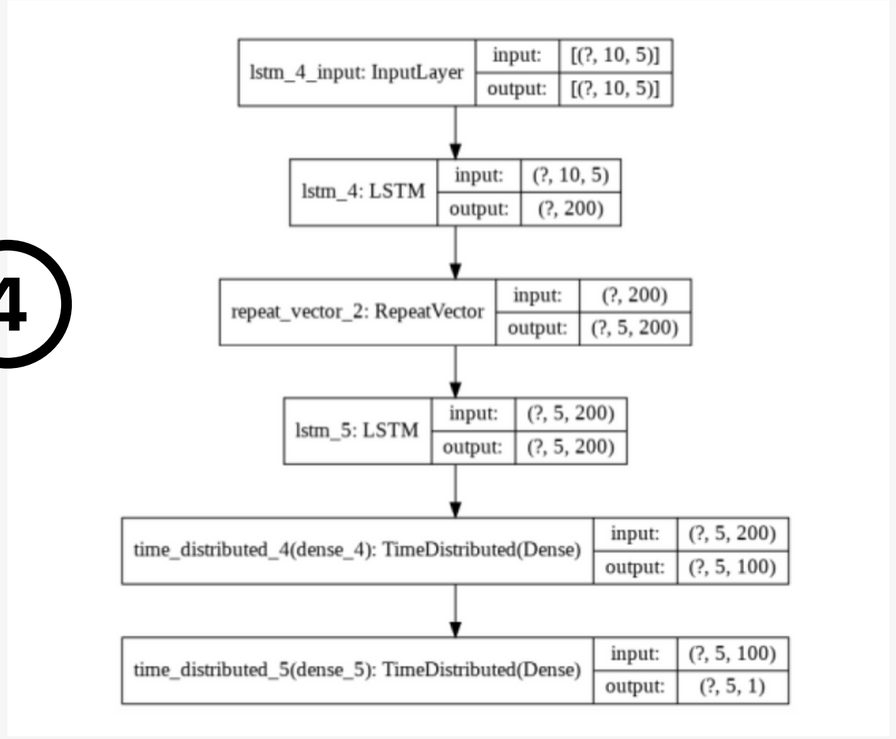
Encoder-decoder LSTM for multi-step forecasting with univariate input data for two weeks

3



Encoder-decoder LSTM for multi-step forecasting with multivariate input data for two weeks

4



Links

OnlyLeaf Link:
https://www.overleaf.com/1369991516jfbbynzgnnfn

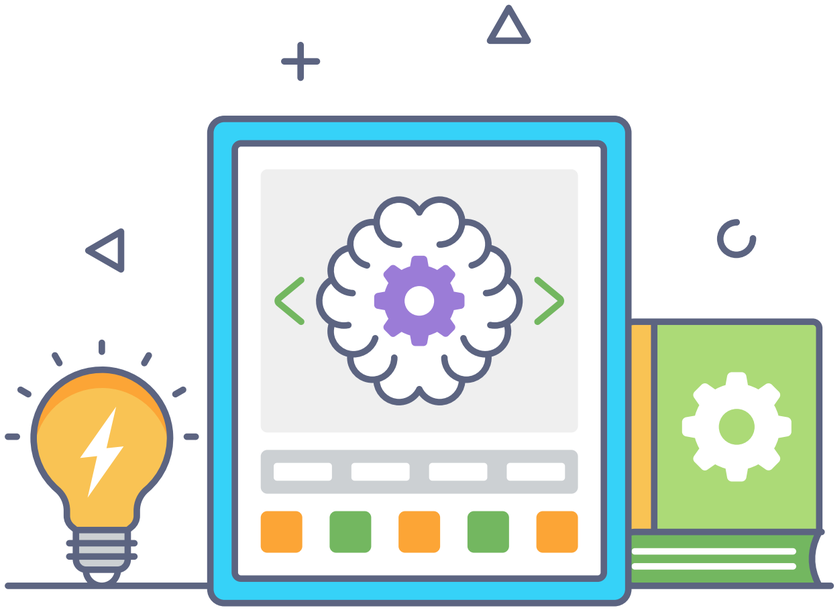
Github Link:
https://github.com/rakshrb1995/Financial-Time-Series-Analysis-with-Machine-Learning.git

MS Teams Folder:
https://deakin365.sharepoint.com/:f:/s/SIT723-FinancialTimeSeriesAnalysiswithMachineLearning/EvrnA2EGws5JsZJ_tPTHamEBPPM_GR2SiPiVsQVHpBjdTw?e=2udk1U

Yahoo 
Finance Data

https://deakin365.sharepoint.com/:x:/s/SIT723-FinancialTimeSeriesAnalysiswithMachineLearning/EY47Al0aEmxHqdUQ9DkbxQ4Bn0KUP2TaQCYYpaTjl9uwRg?e=czBKEO

Plan



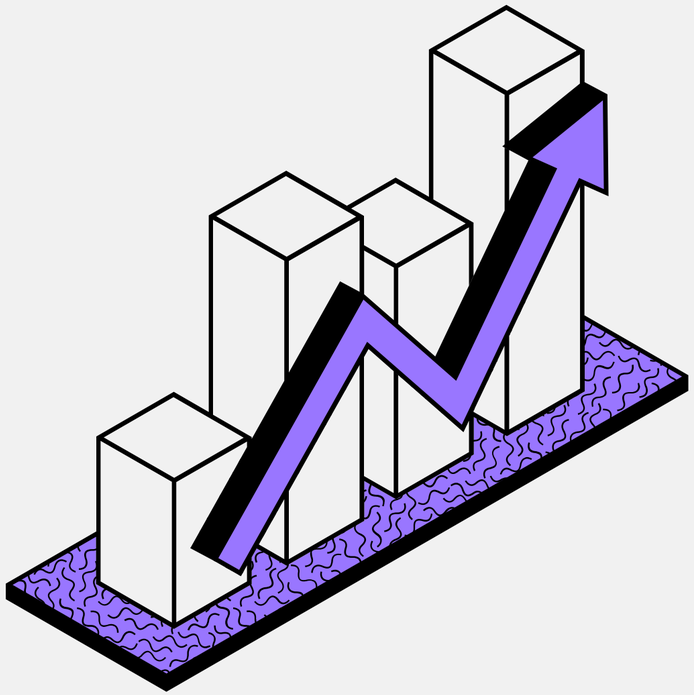
Do a comparative study

Use GAN, Attention models, CNN, ARMIA, RNN - LSTM, Hybrid model integrating ARIMA and LSTM

End of this week

Implement and match the results which is in the paper

Submit literature review draft



Thank you