An Intelligent Predicting Approach Based Long Short-Term Memory Model Using Numerical and Textual Data: The Case of Colombo Stock Exchange

H.N. Kumarasinghe[1], D.M.A.B. Moneravilla[1], I.B.M.R.K.P. Muwanwella[1] and J.B. Ekanayake[2]

[1]*Department of Computer Science and Technology, Uva Wellassa University, Badulla, Sri Lanka*
[2]*Department of Physical Sciences and Technology, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka*

The data forecasting provides a significant guidance for making decisions in many areas especially in stock market today. Due to extremely dynamic and complicated nature of stock markets, price prediction has become a cumbrous challenge. However, there are certain underlying determinants which have a strong influence on the stock market. There are experiments from various areas aiming to take on that challenge and Machine Learning have been the focus of many of them. Nevertheless, many studies used either numerical or textual information, but not both for a single approach. In the present study, a forecasting model was developed to predict the stock prices based on the historical data, investor's activities, macroeconomic variables and news articles. In the process of developing the model, number of factors influencing on stock prices were examined using the ordinary least squares method and technical indicators were identified by reviewing literature. The latest stock data and investor's activities were collected from data library, issued by Colombo Stock Exchange on daily basis for a period of seven years from 2011. Interest rates, exchange rates were used as macroeconomic variables, which were collected from the reports of Central Bank of Sri Lanka. News articles were extracted using a sentimental analysis by analyzing news extracts from most popular news websites. Finally, the prediction model was developed based on recurrent neural network (RNN) and multivariate Long Short-Term Memory (LSTM) approach to predict stock market. The performance of the proposed approach is demonstrated on real-world data of 12 companies listed on Colombo Stock Exchange. The prediction quality of the models is evaluated using MAE, MPE, MAPE, MSE and RMSE. The LSTM and recurrent neural network provided a decent accuracy. The project developed a multivariate prediction model by abolishing the limitation of underutilization of sentiments in price prediction.

Keywords: Stock market prediction, Recurrent Neural Networks (RNNs), Long Short-Term Memory Networks (LSTMs), Multivariate analysis, Sentiment analysis