**INTEGRATED APPROACH FOR STOCK PRICE PREDICTION**

**ABSTRACT**

Stock market analysis has attracted lots of research interests in the literature. Recent studies have shown that Machine learning algorithm achieved better performance than traditional statistical methods. This report introduces a relatively new machine learning technique, support vector machines (SVM) with the integrated approach of Long Short Term Memory concept (LSTM), to the problem in attempt to provide a model with better explanatory power. We used LSTM as a benchmark and obtained prediction accuracy around 99.2% for the integrated aproach of SVM for the TATA Global markets. However, only slight improvement of SVM was observed. Another direction of the research is to improve the interpretability of the three gate model prepared for the deep analysis. We applied recent research results in LSTM model interpretation and obtained relative importance of the input financial variables. Based on these results, we conducted a market comparative predictive analysis on the differences of determining factors in TATA Global company.

**INTRODUCTION**

Stock market prediction has been in the limelight for many decades, but given it's innate complexity and dynamism, it has proven to be a very difficult task. The number of variables and sources of information considered are immense and since there are a lot of locomotive factors affecting the price of the stock, it creates a very effective analysis prediction wise. That makes the task of predicting stock market prices behaviour in the future a very hard one. For many decades, there's been discussions in Science regarding the possibility of such a feat and it's notable in the related literature that most prediction models fail to provide precise prediction in a general sense.

Nonetheless, there is huge amount of research from various disciplines seeking to take on that challenge, presenting a large variety of approaches to attain the supreme goal.

One common approach is to use Machine Learning algorithms to learn from price historic data to predict future prices. This goes in that direction but studying an specific method using recurrent neural networks. Such networks have a short term memory capability and the hypothesis to explore here is that this feature can present gains in terms of results when compared to others more traditional approaches in Machine Learning field.

The algorithm of choice here is the LSTM (Long-Short term memory) network. It's a type of recurrent network that has proved very successful on a number of problems given its capability to distinguish between recent and early examples by giving different weights for each while forgetting memory it considers irrelevant to predict the next output. In that way, it is more capable to handle long sequences of input when compared to other recurrent neural networks that are only able to memorize short sequences.

Historic price data from different stocks from the TATA stock exchange will be used as source of information for the network. Along with this data, a large amount of technical indicators will also be generated to feed the network as features. Upon this dataset the model will be trained, evaluated and will attempt to predict whether the price of a particular stock will go up or not in the next 15 minutes.

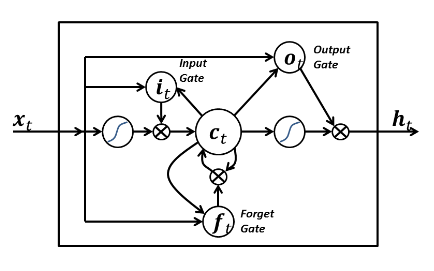
**LSTM has three gates:**

* The input gate: The input gate adds information to the cell state
* The forget gate: It removes the information that is no longer required by the model
* The output gate: Output Gate at LSTM selects the information to be shown as output

The **objective** of this project is to study the applicability of recurrent neural networks, in particular the LSTM networks, on the problem of stocks market prices movements prediction. Assess their performance in terms of accuracy and other metrics through experiments on top of real-life data and analyse if they present any sort of gain in comparison to more traditional machine learning algorithms.

**LITERATURE SURVEY**

When it comes to stock markets, in addition to its inherent complexity and dynamism, there has been a constant debate on the predictability of stock returns. [1] introduced the Efficient-Market hypothesis that defines that the current price of an asset always reflects all previous information available for it instantly. There is also the Random-walk hypothesis [2] which claims that a stock price changes independently of its history, in other words, tomorrow's price will only depend on tomorrow information regardless of today's price. Those two hypotheses determine that there are no means to accurately predict a stock price.



**LSTM MODEL**

**PROPOSED METHOD**

A classification model (Figure 2) was designed based on LSTM networks so as to perform predictions of price movements for a number of stocks. In other words, it attempts to determine if the price of a particular stock will be higher or not than the current price in 15 minutes in the future.

That model is regenerated and trained each trading day on top of historic price data and it is used for performing predictions each 15 minutes using the same model and weights until the end of the day.

Historic price data for particular stocks are gathered in the format of a time series of candles (open, close, high, low and volume) in a granularity of 15 minutes. For this report, data from 2010 to 2017 was collected for stocks that part of the NSE of TATA Global company.

With the data in hand, a log-return transformation is performed as means of normalization as well as to stabilize the mean and variance along the time series. The log-return transformation can be expressed as (1):

**log(pi)−log(pi−1)(1)**

On top of the historic price data, in order to reduce random variation and noise on the pricing series, exponential smoothing was performed through exponentially weighted moving averages (2) as indicated by [19] to cause improvements on the prediction capability.

Also on top of the price data, a set of technical indicators is generated using the TA-Lib1 library. Such indicators are mathematical calculations intended to determine or predict characteristics from stocks based on their historic data. A total of 175 values are generated for each period, and they are intended to represent or predict a very diverse set of characteristics of the stock, like the future price, volume to be traded, the intensity of the current movement tendency, visual graphical patterns, among others.

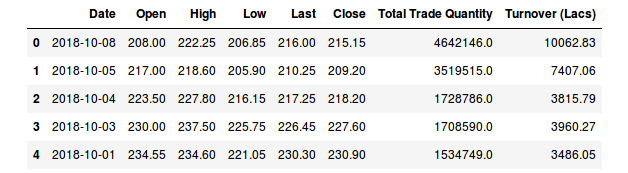
**zi=λx¯i+(1−λ)zi−1(2)**

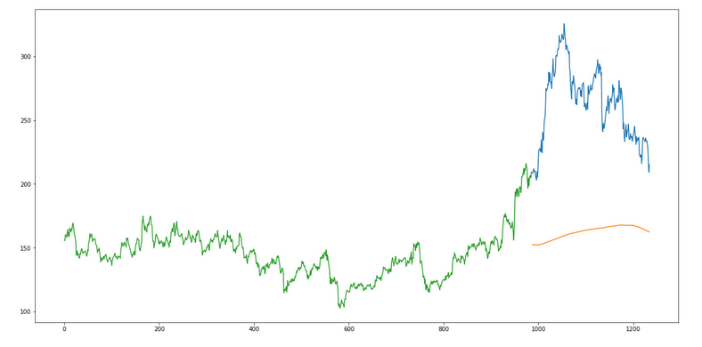
A binary class y is assigned to each entry of the dataset, “1” will indicate that the price will go up on the following time step, and “0” that it won't. Therefore, given that i is the current moment and j is the following, then **j=i+timestep**, and for this project timestep is equivalent to 15 minutes. In case the output is “1”, a “buy” operation will be triggered at i.

For determining the class, the policy to set the value will be based on whether or not the closing price of the next period will be higher than the current one:

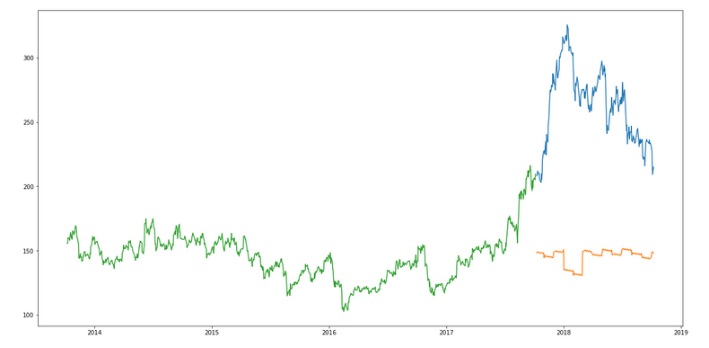
**y={10if closej>closeiif closej≤closei(3)**

The network will take k instances of X as input (Xi−k…,Xi−2,Xi−1,Xi), where X consists in tuples of price data along with technical indicators, and that will be used to predict yj.

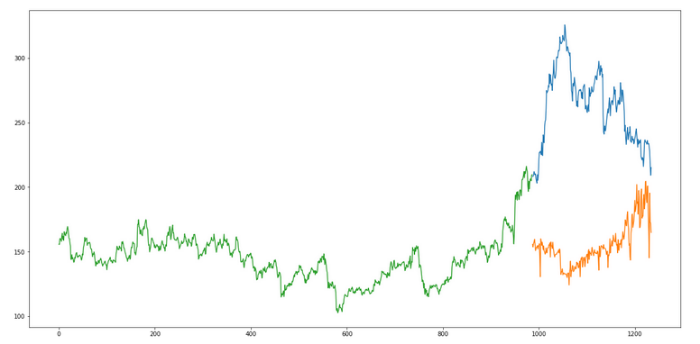




**MODEL 1**



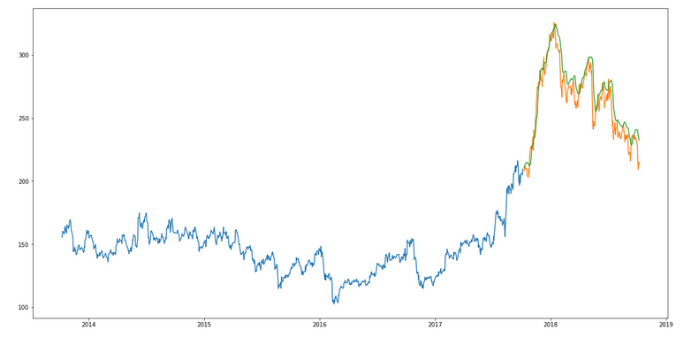
**MODEL 2**



**MODEL 3**

**RESULT**

The LSTM model can be tuned for various parameters such as changing the number of LSTM layers, adding dropout value or increasing the number of epochs. But are the predictions from LSTM enough to identify whether the stock price will increase or decrease? It's a probability.



**CONCLUSION AND FURTHUR WORK**

We can observe that in general the model proposed on this article outperforms the baselines with few exceptions. The outcomes can be considered very promising as it has proven able to predict well compared to other approaches employed today in the literature.

Although the input dimension is very large, the algorithm has demonstrated acceptable capability to learn from it without the need of any dimension reduction technique like feature selection, for example.

When compared to the other machine learning models it has displayed considerable gains in terms of accuracy, but in another hand we believe that variance could be lower and that would contribute for a more reliable model. When it comes to the financial results it's important to note that it was able to keep it positive for all stocks, even though it didn't necessarily had the best results when compared to the baselines.

Another positive aspect is that it had a high return ratio per operation, meaning that it had more success on detecting high variations which is extremely important when account transactions costs and taxes are taken into account. Besides that, when observing the maximum losses it is possible to conclude that the LSTM based model offers less risks when compared to the other strategies.

**REFERENCE**

[1] E. F. Fama and B. G. Malkiel, “Efficient capital markets: Areview of theory and empirical work,”The Journal of Finance,vol. 25, no. 2, pp. 383–417, 1970. [Online]. Available: <http://dx.doi.org/10.1111/j.1540-6261.1970.tb00518.x>

[2] B. G. Malkiel, A Random Walk Down Wall Street. Norton, New York,1973.

[3] A. E. Biondo, A. Pluchino, A. Rapisarda, and D. Helbing, “Are Random Trading Strategies More Successful than Technical Ones?” PLoS ONE, vol. 8, p. e68344, Jul. 2013.

[4] A. Lo and A. MacKinlay, A non-random walk down WallStreet.Princeton, NJ [u.a.]: Princeton Univ. Press, 1999. [Online].Available: <http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+249613484&sourceid=fbwbibsonomy>