AR-LSTM – Adaptative Refined Deep Learning System for Short-term Trend Forecast: A Financial Perspective.

Wilfredo Tovar Hidalgo
School of Information Technology
Carleton University
Ottawa, Canada
wilfredotovarhidalgo@cmail.carleton.ca

ABSTRACT

In the big data era, deep learning and intelligent data mining technique solutions have been applied by researchers in various areas. Forecast and analysis of stock market data have represented an essential role in today's economy, and a significant challenge to the specialist since the market's tendencies are immensely complex, chaotic and are developed within a highly dynamic environment. There are numerous researches from multiple areas intending to take on that challenge, and Machine Learning approaches have been the focus of many of them.

There are multiple models of Machine Learning algorithms been able to obtain competent outcomes doing that class of foresight. This paper proposes the implementation of Long short-term memory (LSTM) networks to predict future trends of a stock, and the novelty of this proposed solution that distinct from previous solutions is that this paper introduced the concept of an adaptative-refined system (AR-LSTM) rather than a sole LSTM model. It was collected data from multiple stock markets such as TSX, SHCOMP, KOSPI 200 and the S&P 500, proposing an adaptative-refined system for trends prediction on stock market prices; It was carried a comprehensive evaluation on several commonly utilized machine learning prototypes, and it is concluded that the proposed AR-LSTM solution approach outperforms preceding models. Additionally, during the research stage from preceding works, gaps were found between investors and researchers who dedicated to the technical domain, and this proposed solution architecture with a comprehensive feature engineering procedure before training the prediction model is pretended to be developed.

CCS CONCEPTS

• Big Data • Deep Learning • Data Mining • Long short-term memory (LSTM) networks.

[*] Adaptative Deep Learning System for Short-term Trend Forecast: A Financial Perspective.
[†]Wilfredo Tovar Hidalgo

KEYWORDS

Deep Learning, Data Mining, Long short-term memory (LSTM) networks, AR-LSTM, Forecast and analysis of the stock market.

I. INTRODUCTION

Prognostications on financial stock markets have been an object of studies, but given its innate complexity, dynamism and chaoticness, it has proven to be a particularly challenging task. The abundance of variables and sources of information held is immense. For decades, there have been studies in Science regarding the possibility of such a feat, and it is notable in the related literature that most prediction models neglect to present accurate predictions in a broad sense. Nevertheless, there is a tremendous number of studies from various disciplines seeking to take on that challenge, presenting a large variety of approaches to reach that goal.

Since the forecast of the financial stock market has drawn attention from industry to academia as in [1] [2]. Various machine learning algorithms, such as genetic algorithms, support vector machines, neural networks, among others, have been in use to predict the market variances. Recurrent neural networks (RNNs) hold a robust model for treating sequential data such as sound, time-series data or written natural language. Some designs of RNNs were used to predict the stock market as in [3] [4].

The traditional approach is to use Machine Learning algorithms to learn from historical price data to predict projected values. This proposal intends to maintain that direction but studying a specific method employing recurrent neural networks. Such networks have a short-long term memory capabilities and the hypothesis to explore feature pursuing gains in terms of results when is compared to other more traditional approaches in Machine Learning field, by introducing the concept of an adaptative-refined system (AR-LSTM) rather than a sole LSTM model, since, during the research stage from preceding works, gaps were found between investors and researchers who dedicated to the technical domain, and this proposed solution architecture with a comprehensive feature engineering procedure pretended to train the prediction model to outperform the traditional LSTM approach.

In this project, were used multiple cleaned datasets organized and formed independently. However, the data is an open-sourced data from multiple markets obtained at the Yahoo Finances platform; this paper illustrates the data collection details in section IV.

II. PROBLEM STATEMENT

This research project determines three main research questions fundamental on an ample preceding literature review and suggests a comprehensive solution supported by a full evaluation that intends to solve the following research interrogations.

1. How engineering features benefit the model's prediction accuracy?

From the plentiful previous publications evaluate, we can assume that stock price data are embedded with a high level of statistical dissonance, and there are relationships among features, which makes the price prediction particularly complex to predict. That is the principal cause for most of the previous relate works introduced the feature engineering part as an optimization module.

2. What is the best algorithm for prediction short term price trend?

Assume from previous works, researchers are pursuing the exact price prediction. While this will be intended to crumble the problem into predicting the trend and then an exact number, this paper will concentrate on the first step. Therefore, this objective will be converted to solve a binary classification problem, meanwhile discovering an effective way to eliminate the negative effect brought by the high level of statistical noise. This approach will decompose the complex problem into sub-problems that produce fewer dependencies and solve them one by one, then compile the resolutions into an ensemble model as an aiding system for investing behavior reference.

3. How do innovations from economic-financial field privileges prediction model design?

Distinct from previous works, besides the standard evaluation of data models such as the training costs and rates, this paper pretends to evaluate the effectiveness of recently added features that can be extracted from the financial domain. It will include some features from the financial domain. While is obtained some specific findings from previous research works, and the related raw data needs to be processed into usable features. After extracted related features from the financial domain. This paper will combine the features with other standard technical indices for voting the features with high impact.

There are numerous features said to be valid from the financial area. Thus, how to accurately transform the conclusions from the financial field to a data processing module of the proposed system design is a covered research question that must be facing within this research.

III. SURVEY OF RELATED WORK

Due to its complexity and dynamism, there has been a constant dispute on the predictive performance of several stock returns predictors. In regard to computational intelligence, there are several studies assessing various methods in order to accomplish accurate predictions on the stock market. K. Kim in [5] Uses statistical learning by using algorithms like Support Vector Machines (SVM), Estrada [6] Proposed a new approach based on collective intelligence, including neural networks, component modelling, textual analysis based on news data.

There are effective works related to deep learning in stock markets; there are some examples like Sharang Et al. [7] published a study where is performed on the application of a Deep Belief Network (DBN), which is composed of stacked Restricted Boltzmann Machines, coupled to a Multi-level Perceptron (MLP) and using long-range log returns from stock prices to predict above-median returns. Heaton Et al. [8] implemented DBN, using price history in addition to technical indicators as input, in a similar approach to this project. Both of those works present improved results matched to their baselines, as well as in Greff Et al. [9] presented a survey in deep learning methods applied to finance is done and their improvements discussed.

Lei in [10] utilized Wavelet Neural Network (WNN) to prognosticate stock price trends, this approach also included Rough Set (RS) for attribute reduction as an optimization, it was employed to reduce the stock price trend characteristic dimensions and to define the composition of Wavelet Neural Network. The prototype evaluation was proved on different stock market indices, and the result was satisfying with generalization. It demonstrates that Rough Set would effectively decrease the computational complexity. However, the author only emphasized the parameter adaptation in the discussion part while it did not define the vulnerability of the model itself. In consequence, during the literature review, it was found that the evaluations were conducted on indices, and performance varies if it is implemented on specific stocks. The features table and calculation formula are worth taking as a reference.

McNally et al. in [11] Presented a work based on Recurrent neural network (RNN) and LSTM to prognosticating cryptocurrencies by applying the Boruta algorithm for the feature engineering part, which operates comparable to the random forest classifier. They also employed Bayesian optimization to select LSTM parameters. The principal dilemma of their approach is overfitting. The research problem of predicting the cryptocurrencies price course has remarkable relations with stock prediction. There are unknown characteristics and noises inserted in the price data, and treat the research problem as a time sequence problem.

Fischer and Krauss in [12] applied long short-term memory (LSTM) successfully on financial stock market prediction. Their dataset was compounded by the S&P 500 index, where the lists were consolidated into a binary matrix to eliminate survivor bias. The authors also applied effectively an optimizer called "RMSprop, "which is a mini-batch variant of rprop. This paper offers a robust notion of times series predictions, but it is not suitable for PRICE Trend Forecast. The principal strength of this research is that the designers adopted the most advanced deep learning technique by 2018 to offer financial market prognostications. However, they relied on the LSTM system, causing a misconception of its practical implementation, produced by a lack of background and understanding of the financial domain. Though the LSTM outperformed the conventional DNN and logistic regression algorithms, the author did not consider the effort and application to train an LSTM with long time dependencies. Consequently, Their approach showed that LSTM is suitable for financial time series prediction tasks that are a different domain than Short-term PRICE Trend Forecast, which is the aim of the proposed approach on this paper.

A. Gap Analysis

This section illustrates the gaps found from the information and comparison of contents of related researches.

The gaps found among the research articles in the finance field and technical domain are mainly on data preprocessing. Technical-Scientific research papers tend to focus on building the models more effectively, but occasionally, they do not consider the applicability of their designed system and models. In research papers, there is a process of selecting the features that will be covered and are examined and mentioned from preceding works and go through the feature selection algorithms. While in the financial domain, the researchers show more interest in behaviour analysis, since they may affect the stock performance. During their studies, they frequently conduct a full statistical analysis based on a particular dataset and decide new features rather than performing feature choosing.

From those findings, It may be inferred that the features or behaviours being focused on the financial domain are rarely being investigated in the technical domain, and the financial domain research paper also hardly introduced the machine learning or deep learning algorithms to train their model without implementing data preprocessing method as investors do. So, it came with a consideration that by selecting new features from the data, then combine the features with existed standard technical indices would benefit the actual and proved models.

IV. PROCEDURE OF THE STUDY

A. Datasets Preparation and Description

The experiments used historical closing price data from the following stock markets indices, i.e., TSX, SHCOMP, KOSPI 200 and the S&P 500, and included all stocks in each index from 31 January 2012 to 31 July 2018 on a working day basis. These data were obtained from the Yahoo financial website (https://finance.yahoo.com/). This dataset consists of 2954 stocks. It was collected everyday price data, stock ID, reopening history, and the top 5 main shareholders. The data used were from 31 January 2012 to 31 July 2018, since exercising with more up-to-date data would help the study result.

B. Dataset Structure Design.

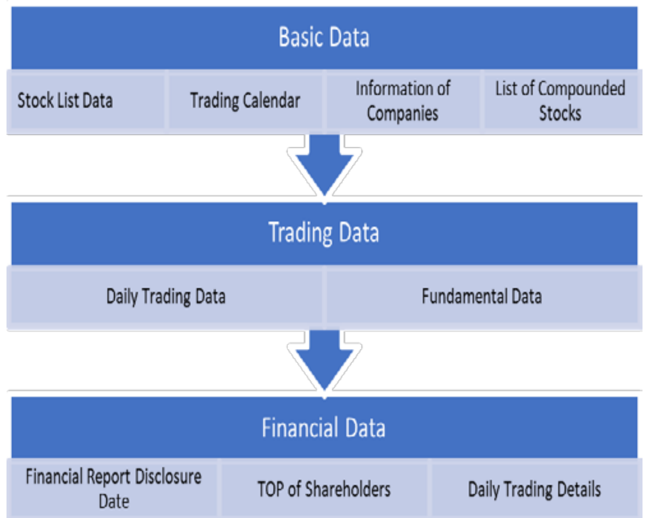

Figure 1. Dataset Structure Design.

Figure 1 shows the total of the data on the dataset tables. There are three classes of data in this dataset: basic data, trading data, finance data.

The totality of the data table can be associated with a track denominated "S-ID. Or Stock-ID" It is an individual stock identifier recorded in the commodities markets.

Basic Data is the essential information that the users or researchers might require during the exploration process of the data. It is compounded for Stock-List Data, Trading Schedules, Primary Information of Listed Corporations, etc.

Trading data is the value and trade-related data for a financial instrument published by a trading regulator or market.

Finance data, the information considered on this case was based on the income statement and balance sheet of every Stock-ID.

C. Technical Background & Indices [13]

In this section, the most regularly employed technical indices are illustrated.

1) Stochastic indicator K.

The n-day stochastic indicator K is defined as:

$$K\_n_i = \frac{2}{3} \leftrightarrow K\_n_{i-1} + \frac{1}{3} \leftrightarrow \frac{CP_i - LP\_n_i}{HP\_n_i - LP\_n_i} \leftrightarrow 100$$

2) Stochastic indicator D.

The n-day stochastic indicator D

$$D\_n_i = \frac{2}{3} \times D\_n_{(i-1)} + \frac{1}{3} \times K\_n_i$$

Where K_ni is the n-day stochastic indicator K of day i.

3) Williams overbought/oversold index

The n-day Williams overbought/oversold index is a momentum indicator that measures

overbought and oversold levels.

$$WMS\%R\_n_i = \frac{HP\_n_i - CP\_i}{HP\_n_i - LP\_n_i}$$

4) Commodity channel index

The commodity channel index is used to identify cyclical turns in commodities.

It is define the typical price as the formula below:

$$TP_i = \frac{HP_i + LP_i + CP_i}{3}$$

Then we calculate the n-day simple moving average of the typical price:

$$SMATP\_n_i = \frac{\Sigma_{j=i-n-1}^{i} TP_j}{n}$$

And the n-day mean deviation is noted by MD_n:

$$MD\_n_i = \frac{\Sigma_{j=i-n-1}^{i} |TP_j - SMATP\_n_i|}{n}$$

The n-day commodity channel index is calculated as:

$$CCI\_n_i = \frac{TP_i - SMATP\_n_i}{0.015 \times MD\_n_i}$$

5) Relative strength index, The relative strength index is a momentum oscillator that compares the magnitude of recent gains to the magnitude of recent losses.

$$G_i = \begin{cases} CP_i - CP_{i-1} & \text{if } CP_i > CP_{i-1} \\ 0 & \text{otherwise} \end{cases}$$

6) Moving average convergence/divergence

The moving average convergence/divergence is a momentum indicator that shows the relationship between two moving averages.

First step is to calculate the demand index (DI):

$$DI_i = (HP_i + LP_i + 2 \times CP_i) / 4$$

We also need to calculate the 12-day and 26-day exponential moving average:

$$EMA\_12_i = \frac{11}{13} \times EMA\_12_{i-1} + \frac{2}{13} \times DI_i$$

And

$$EMA\_26_i = \frac{25}{27} \times EMA\_26_{i-1} + \frac{2}{27} \times DI_i$$

Hence, we use DIFi to indicate the difference between EMA_12 and EMA_26:

$$DIF_i = EMA\_12_i - EMA\_26_i$$

The MACDi is calculated as below:

$$MACD_i = \frac{8}{10} \times MACD_{i-1} + \frac{2}{10} \times DIF_i$$

7) 10-day moving average

The 10-day moving average is the mean price of the futures over the most recent 10 days and is calculated by:

$$MA\_10_i = \frac{\Sigma_{j=i-9}^{i} CP_j}{10}$$

8)Momentum

Momentum measures change in stock price over last n days.

$$MTM_i = \frac{CP_i}{CP_{i-n}} \times 100$$

9) Rate of Change

The n-day rate of change measures the percent changes of the current price relative to the price of n days ago and is calculated by:

$$ROC\_n_i = \frac{CP_i - CP_{i-n}}{CP_{i-n}} \times 100$$

10) Psychological line

The psychological line is a volatility indicator based on the number of time intervals that the market was up during the preceding period and is calculated by:

$$PSY\_n_i = \frac{TDU\_n_i}{n} \times 100\%$$

The TDU_ni is the total number of days that has price rises in previous n days.

11) AR

n-day A ratio means the n-day buying/selling momentum indicator which is calculated as:

$$AR\_n_i = \frac{\Sigma_{j=i-n-1}^{i}(HP_j - OP_j)}{\Sigma_{j=i-n-1}^{i}(OP_j - LP_j)}$$

12) BR

n-day B ratio means the n-day buying/selling willingness indicator and is defined as:

$$BR\_n_i = \frac{\Sigma_{j=i-n-1}^{i}(HP_j - CP_{j-1})}{\Sigma_{j=i-n-1}^{i}(CP_{j-1} - LP_j)}$$

13) Volume ratio

The n-day volume ratio is calculated by:

$$VR\_n_i = \frac{TVU\_n_i - TVF\_n_i / 2}{TVD\_n_i - TVF\_n_i / 2} \times 100\%$$

Where the TVU represents the total trade volumes of stock price rising, and TVD is the total trade volumes of stock prices falling, TVF represents the total trade volumes of stock prices holding in previous n days.

14) Accumulation/distribution oscillator

$$AD_i = \frac{HP_i - CP_{i-1}}{HP_i - LP_i}$$

15) 5-day bias

The 5-day bias is the deviation between the closing price and the 5-day moving average MA_5

$$BIAS\_5_i = \frac{CP_i - MA\_5_i}{MA\_5_i}$$

V. PROPOSED SOLUTION

The proposed scheme solution consists of three parts. First is the feature choice section; it ensures the chosen features are all effective. Second, the data and perform the dimensionality reduction, and the Third, the building of a forecast model that targets stocks, which is the principal contribution of the work.

1. Selection:

There are multiple methods to distinguish various classes of stocks according to the preferences of the investors; some of them prefer lengthy-term investment, while others present more attention in short term investment, also known as speculative investments. It is normal to observe in the stock report of a company that is reporting an ordinary performance; the stock price increasing drastically; this is one of the common events that illustrate that the stock price prognostication has no fixed rule, thus find effective features before train the data model is required. This proposal is focused on the short term price trend prediction, which makes this labour even more complex.

It was determined the data with no labels; the initial action is to label the data. Identification should be added to the price trend by examining the actual closing price with the closing price of n trading days ago, the range of n is between 1 to 10 considering that this research is focused on the short term predictions. If the price trend increase, it has to be identified as 1 or mark as 0 in the opposing scenario. For purposes of optimization, the indices were employed from the indices of n−1th day to predict the price trend of the n th day.

2. Dimensionality:

It is required to filter the high-quality stocks to ensure the best execution of the prognostication model; the data first has to be evaluated as a primordial step. The recursive feature elimination (RFE) was implemented to guarantee that all the chosen characteristics are useful since the raw data contains a vast number of features. If we include all the features into consideration, it will drastically increase the computational complexity and additionally produce side consequences for further research if we would like to perform unsupervised learning.

It was noticed that most of the previous works in the technical domain were examining all the stocks, while in the financial field, researchers favour the analysis of the specific scenario of investment for its practicality and look for results, to fill the gap between the two domains, It was decided to apply a feature extension based on the findings collected from the financial domain before the start the RFE procedure.

3. Forecasting:

The plan of this approach is to model the data into time series to short-time forecasting instead of the realization of the time series at all; in consequence, the larger number of the characteristics, the more complex the training procedure will be. So, to tackle this problem, the dimensionality reduction was implemented by using Randomized PCA at the origin of our suggested solution architecture, providing novelty of Refined Adaptive features to the traditional LSTM forecasting.

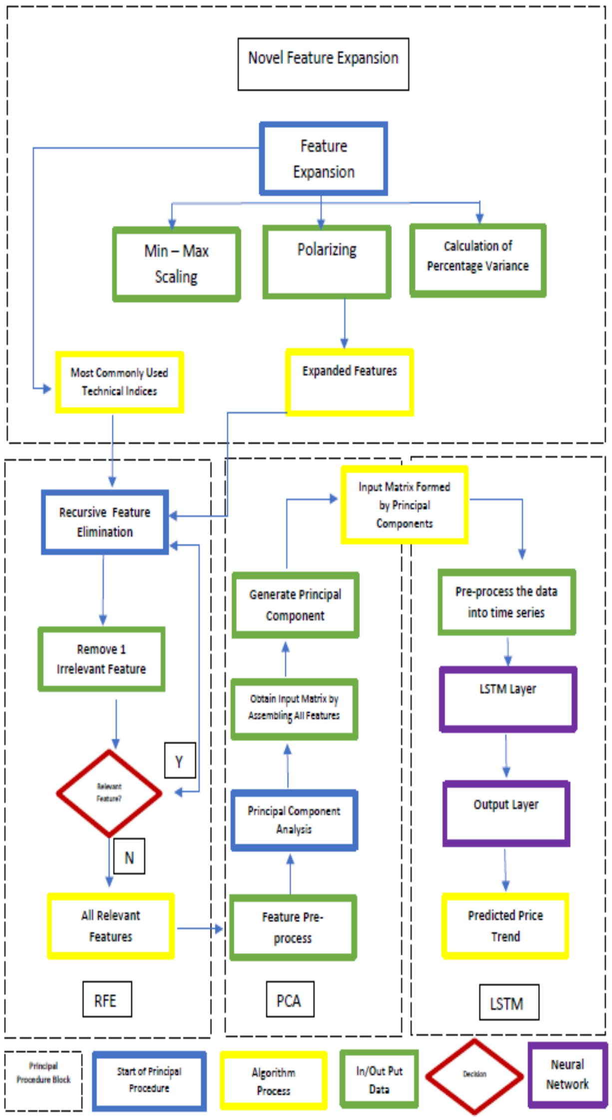

Figure 2. Novel Technical Design.

- **Technical Design Implementation**

The full stages process is illustrated in Figure 2.

The components were split by principal procedures, where each procedure consists of algorithm steps. Algorithmic features will be discussed later in the paper.

Our proposed select technical indices which are the most studied and proved to be effective i features from the extended feature set. Next step, is to feed the data with i chose features into PCA algorithm to degrade the dimension into j features, after it obtained the best combination of i and j, the final step consists of the data into finalized the feature set and feed them into the LSTM model to perceive the price trend prediction result.

The novelty of this proposed solution is that it will not exclusively employ the technical method on raw data but additionally implement the feature extension that used among stock market investors.

A. Feature extension.

One of the principal procedures in Figure 2 is the Feature extension. In this block, the input information is the most generally employed technical indices concluded from related works. The three feature extension methods are max-min scaling, polarizing, and calculating fluctuation percentage.

B. Recursive Feature Elimination.

In this step, it is studied and applied the most efficient i features by practicing the Recursive Feature Elimination (RFE) algorithm, its estimate all the characteristics/features by two properties, coefficient and importance. It also restricts the remove features from the pool by 1, which implies it will eliminate one characteristic at each round and hold all the relevant features. Finally, the output of RFE block will be the input of the next step, -> PCA.

C. Principal Component Analysis

It will reflect the principal element extraction result from RFE. Therefore, before filling the data within the PCA algorithm, a feature preprocessing is necessary.

Following finishing characteristic pre-processing, the next action is to serve the treated data with chosen i features into PCA algorithm to decrease the feature matrix scale into j features. This action is to employ as many useful features as practicable and meanwhile reduce the computational complexity of training the model. This proposed solution also works on the evaluation of the balanced pair of i and j, which has comparatively more solid prognostication accuracy; at the time, it cuts the computational consumption.

After the PCA action, the system will capture a reshaped matrix with j columns.

D. Long ShortTerm Memory

PCA decreased the measure of input data, while the data pre-processing is necessary before serving the data into the LSTM layer.

The reason for preprocessing is due that in LSTM, the input matrix formed by principal elements has no time steps. However, one critical parameters of training an LSTM is the number of time steps. Consequently, the matrix has to be modelled into corresponding time steps as for both training and testing datasets.

The last action is to fill the training data into LSTM and assess the execution employing testing data. Two layers form the LSTM structure. The input dimension is defined by j after PCA algorithm. The initial layer is the input LSTM layer, and the second layer is the output layer. *The final output will be 0 or 1 indicates if the stock* *price trend forecast results are proceeding down or going up, as supporting advice for the investors to execute the next investment decision.*

- ## Algorithm Creation

The algorithmic detail was generated, sequentially, the initial algorithm is the crossed feature engineering element for developing high-quality training and examining data. It corresponds to the Feature extension, RFE and PCA blocks in Figure 2. while the following algorithm is the LSTM system block, including time series data, and pre processing, NN, training, constructing and testing.

Algorithm 1: Short term Price Trend Forecasting - Crossed Feature Engineering employing PCA / FE / RFE

The function FE() is relating to the feature extension block. In terms of the feature continuation procedure, It was employed three distinct processing operations to translate the conclusions from the financial field to a technical module in the system scheme.

The function RFE() in the algorithm applies to recursive characteristic elimination. It performs the practice data scale decrease.

PCA() refers to Principal Component Analysis. As explained in [14], PCA is an algorithm that often used in feature engineering; it will convert the fundamental variables into new variables with the largest data held. For the training data matrix scale reduction, It was employed the Randomization of Principal Component Analysis (PCA), where Economic rates of a recorded business are utilized to exhibit the germination capacity, gaining ability, stability, among others characteristics. A determined Business rate is compounded by a set of technical indices; a new data column into the data matrix is aggregated at the time of adding a new technical index.

Algorithm 1

```
fe_array = fe[i, 5]

function FE(f)
    for m in range[0, 5] do
        if fe_array[i − 1, m] == 1 then
            Feature expansion method
        end if
    end for
     return df_X_FE    ▷ (df_X_FE is the processed data frame after feature expansion)
end function

function RFE(df)                    ▷ (Recursive feature elimination function)
    Tune the model on the training set with all features
    Calculate model performance with testing samples
    Ranking the weight of different features
    for Each subset do
        Retain i most weighted features
        Tune the model on the training set with all features
        Calculate model performance with testing samples
    end for
    Calculate performance profile over testing samples
    Estimating the features by final testing dataset
    Fit the final model based on selected features using the original training set
     return df_X_RFE   ▷ (df_X_RFE is the processed data frame after RFE algorithm)
end function

function PCA(df)            ▷ (Leverage optimization algorithm PCA to reduce dimension from i to j)
    n_components=j, whiten=False, copy=True, batch_size=250
     return df_X_PCA  ▷ (df_X_PCA is the optimized data frame after PCA algorithm)
end function

function MAIN( )                                        ▷ (Main function)
    FE(f[i])
    DATAPARTITION(df_all, method = resampling)
    RFE(df_X_FE)
    PCA(df_X_RFE)
     return df_X_PCA
end function
```

Algorithm 2: Price Trend Prognostication Design Using LSTM.

It employed an LSTM model and attached a regeneration method for the stock price dataset. The function TIMESSCONVERSION() transforms the main elements within the time series by changing the input data structure according to the quantity of time steps, for example, the term dimension in this project. The prepared dataset consists of the input string and forecast series. In this project, the parameter of lag is 1, since the model is identifying the design of features variation daily. The NTIMESTEPS is diversified from 1 trading day to 3 trading days. The NN composition scheme, optimizer choice and other parameters are represented in function MODELCOMPILE().

Algorithm 2

```
function    TimesSConversion  (df, term_length, lag)

    cols = list()
    for i in range (term_length, 0, -1) do
        shift df by i
        append shifted df to cols
    end for
    for i in range (0, lag) do
        shift df by -i
        append shifted df to cols
    end for
    df_X_TS = concat(cols, axis = 1)
     return df_X_TS
end function

function MODELCOMPILE(j)                 ▷ Define NN structure and compile
    Stack_method = Sequential()
    Layer_1 = LSTM(50, input_shape=(train_X.shape[1], train_X.shape[2]))
    Layer_2 = Dense(1)
    Loss_Function=mae
    Optimizer=adam
    Metrics=f1, metrics.binary_accuracy, metrics.mean_squared_error, metrics.mean_absolute_error
     return LSTMmodel
end function

function MAIN( )                                        ▷ Main Function
    TIMESERIESCONVERSION(df_X_PCA, N_TIME_STEPS, LAG)
    DATAPARTITION(df_X_TS, method = resampling)
    MODELCOMPILE(j)
    FITMODEL(X, y, epochs=50, batch_size=3000)
    EVALUATEMODEL(X_test, y_test)
end function
```

The codes referring to the proposed approach will be publicly available, at GITHUB under the user name WilfredoTovar.

VI. RESULTS AND EVALUATION

The evaluation of the algorithm was produced on a Hp laptop with 2.2 GHz Intel Core i7 processor, embedded 16 GB RAM memory.

Randomly was chosen one-thirds of the stock data for RFE training and notes the dataset as DS_train_f. The estimator of the RFE algorithm is SVR with a linear kernel. We place the 54 features by polling and prepare 30 useful features to process it using the PCA algorithm to perform dimension reduction and decrease the features into 20 principal components. The residue of stock data forms the testing dataset DS_test_f to validate the effectiveness of essential elements that were extracted from selected features.

It was tested the RFE algorithm on a series of short term from one to six days, to estimate how generally related technical indices corresponded to the price course.

- Throughout the examination, it was observed that varying range of terms has a distinct level of sensitiveness to the very indices set.
- It was obtained the close value of the first dealing date and contrast it with the close price of the n_th trading date. Since this approach predicts the price course, it was not considered the term lengths if the cross-validation score is below 0.5. They are n = {2, 3,5}, which shows that the price trend forecast for every different day using the indices set is likely to be more reliable than in more prolonged periods.

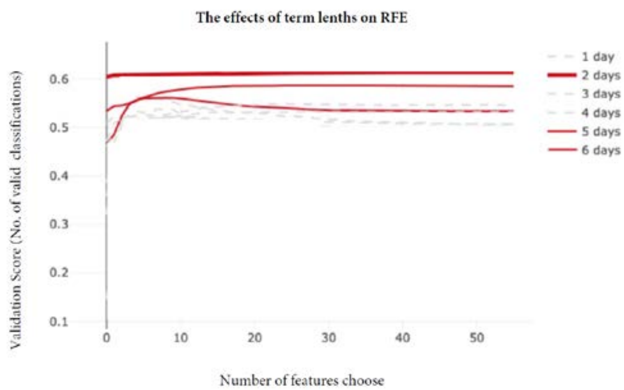

Figure 3. The effects of term lengths on RFE.

Figure 4. Shows the performance of the proposed approach in a range of 6 days of short-time forecast, where it demonstrates a correct performance by approaching market trends accurately, even in high volatility scenarios such as abrupt falls or market stoppages.

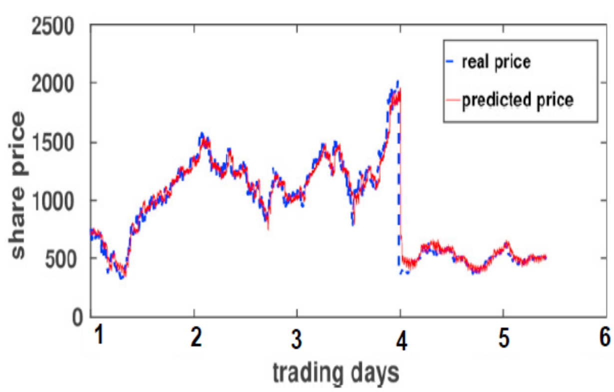

Figure 4. Short-time prediction assertiveness.

In Figure 5. The performance of the traditional LSTM approach in short-term predictions is illustrated on the left of the graph. It can be observed that the traditional approach is capable of identifying price trends effectively. However, it demonstrates a high volatility or dissonance to achieve predict PRICE accurately in short times, which for speculative transactions makes it ineffective since they required a higher degree of effectiveness on price prediction, not only trends.

On the right side of the graph, the performance of the proposed AR-LSTM approach can be observed, and it is shown that outperform the results of the traditional approach since it effectively predicts market trends and additionally demonstrates a lower degree of dissonance with respect to to the price prediction. Since it possesses more accurate results on price prediction makes it more useful for the evaluation of applications in speculative market transactions such as the daily currency market or commodities.

The results demonstrate that the proposed AR-LSTM approach outperforms in performance for short-term price predictions compared to the traditional LSTM.

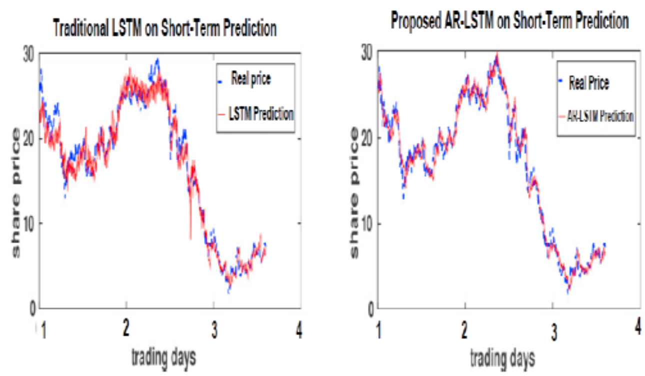

Figure 5. Short-time prediction assertiveness Traditional LSTM vs Proposed AR-LSTM.

.

VII. FUTURE WORK AND RESEARCH OPPORTUNITY

The research possibility is ample considering it was managed a mixture of data; Firstly, the regular trading data is available, researchers may employ the price information to determine largest of the technical indices, also they can be modelled with necessary prices in a determine range of time to apply time series and obtain the price or trend prediction also known as "Time Series Predictions." Secondly, the value and indices may be utilized as characteristics, and joined with other data assembled in the dataset has the potential used as characteristics that may be employed for data mining purposes. Finally, maybe leverage the news scouring on media to observe the process and effects of posting and its effect on the stock market price, which eventually explore a real-time sentiment analysis system.

On this paper have produced a satisfactory prognostication of price result from the proposed model. However, this research project still possesses considerable potential in future research. First of all, the objective of building the model to produce short term price trend prediction is to complete the very first step of stock market price prediction. With a reliable trend prediction, we can perform the price prediction in a more reliable way for different magnitudes of times.

VIII. CONCLUSION

This research paper offered multiple contributions. In this project, after performing a comprehensive literature review in the financial domain, it was gathered the data portions that are useful for stock market analysis, then it was assembled, clean-up and structure it market data for future research. Second, was obtained two subsets of additions: i) examining into the techniques frequently employed by real-world investors, it was developed a new algorithm element and defined it as Adaptative Refined extension, then was employed FE, RFE AND PCA algorithms to build a feature engineering method which is both effective and efficient. It was determined that the AR-LSTM prognostication model, produced an important high prediction accuracy that outperforms the traditional model for short-time price prediction. In addition to the contribution above-mentioned, this work carries a comprehensive evaluation by comparing one of the most frequently used machine learning models with our proposed AR-LSTM; we conclude multiple heuristic findings that possess the potential to be future research questions in both technical and financial research domains.

REFERENCES

[1] Krollner, Bjoern, Bruce Vanstone, and Gavin Finnie. "Financial time series forecasting with machine learning techniques: A survey." (2010).

[2] Agrawal, J. G., V. S. Chourasia, and A. K. Mittra. "State-of-the-art in stock prediction techniques." International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering 2.4 (2013): 1360-1366.

[3] Saad, Emad W., Danil V. Prokhorov, and Donald C. Wunsch. "Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks." Neural Networks, IEEE Transactions on 9.6 (1998): 1456-1470.

[4] Rather, Akhter Mohiuddin, Arun Agarwal, and V. N. Sastry. "Recurrent neural network and a hybrid model for prediction of stock returns." Expert Systems with Applications 42.6 (2015): 3234-3241.

[5] K. Kim, "Financial time series forecasting using support vector machines," Neurocomputing, vol. 55, no. 1–2, pp. 307 – 319, 2003, support Vector Machines. [Online]. Available: http://www.sciencedirect. com/science/article/pii/S0925231203003722.

[6] B. Batres-Estrada, "Deep learning for multivariate financial time series," 2015.

[7] A. Sharang and C. Rao, "Using machine learning for medium frequency derivative portfolio trading," CoRR, vol. abs/1512.06228, 2015. [Online]. Available: http://arxiv.org/abs/1512.06228

[8] J. B. Heaton, N. G. Polson, and J. H. Witte, "Deep learning in finance," CoRR, vol. abs/1602.06561, 2016. [Online]. Available: http://arxiv.org/abs/1602.06561

[9] K. Greff, R. K. Srivastava, J. Koutn'ık, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," arXiv preprint arXiv:1503.04069, 2015.

[10] Lei, L. (2018). Wavelet neural network prediction method of stock price trend based on rough set attribute reduction. Applied Soft Computing, 62, 923-932.

[11] McNally, S., Roche, J., & Caton, S. (2018, March). Predicting the price of bitcoin using machine learning. In 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP) (pp. 339-343). IEEE.

[12] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654-669.

[13] Gorgulho, A., Gorgulho, A. M., Neves, R. F., & Horta, N. (2012). Intelligent financial portfolio composition based on evolutionary computation strategies. Springer Science & Business Media.

[14] Pang, X., Zhou, Y., Wang, P., Lin, W., & Chang, V. (2018). An innovative neural network approach for stock market prediction. The Journal of Supercomputing, 1-21.

[15] Nguyen, D. H. D., Tran, L. P., & Nguyen, V. (2019, November). Predicting Stock Prices Using Dynamic LSTM Models. In International Conference on Applied Informatics (pp. 199-212). Springer, Cham.

[16] Liu, Y. (2019). Novel volatility forecasting using deep learning–long short term memory recurrent neural networks. Expert Systems with Applications, 132.

[17] Shen, J. (2019). Short-term Stock Market Price Trend Prediction Using a Customized Deep Learning System.

[18] Yahoo. (2018). NIFTY 50 (NSEI)/S&P BSE SENSEX (BSESN). Retrieved April, 2020, from https://in.finance.yahoo.com

[19] Yahoo. (2019). S&P 500 Stock Data. Retrieved April, 2020, from https://finance.yahoo.com/quote/%5EGSPC/history?p=%5EGSPC