Turkish Journal of Computer and Mathematics Education *Vol.12 No.6 (2021), 4529-4534*
Research Article

Time Series Model for Stock Market Prediction Utilising Prophet

Dr. Lilly Sheeba. S, Neha Gupta, Anirudh Ragavender R M, D. Divya
Chennai

Abstract: The importance of predicting the rise and fall of the stock market is indisputable, as it helps investors make wise decisions in terms of buying and selling stocks; however, underlying nuances of the stock market make it difficult to build accurate prediction models. Time is an important facet in determining stock trends, which unfortunately is always neglected. Our approach highlights the significance of time to improve the accuracy of the prediction, which is done by utilising the prophet library. In order to improve the accuracy, the dataset[2] was thoroughly investigated and visualised by constructing numerous graphs which shed light on how time has impacted the stock prices. The prophet library defines three hyperparameters namely seasonality, trend, and holidays. Each of these parameters elucidate the gravity held by time in stock market prediction. We have employed the dataset [2] of a multinational financial services company called the Banco Santander, S.A., this dataset[2] contains the stock prices of the Santander Group, which is Located in Spain. Moreover, it is the 16th largest banking institution in the world. It was hypothesised that data which is nearer to the data to be predicted held much more significance as compared to historic data and this hypothesis was validated, indicating the importance of time based data for an accurate prediction. The month of May saw a drastic increase in the stock price of our dataset[2] and a slight increase on Thursdays.

Keywords: Banco Santander Dataset[2], Prophet Library, Stock Market, Time Series.

1. Introduction

The process of determining the future value of an entity's stock or any other financial instrument traded on an exchange, accounts to stock market prediction. It has grown to be an integral component in investing with the rise of compute power in cloud technologies. Stock market prediction has been a buzzword but the results of it were not up to the mark. To conquer this challenge multiple facets of the stock market have been analysed but the highly volatile nature of the stock market makes it difficult to predict. The volatility associated with the stock market has plagued it since its conception. The cause of this inconsistency can be attributed to its heavy dependence on time. Unfortunately, the element of time is always considered to be the same as any other attribute. In order to elaborate on the gravity associated with time this paper uses it as an integral constraint. This is done by exploiting time related data by employing the prophet library. The basic idea behind the employment of prophet is to ascertain the interplay between time and the stock market. By doing this we can reduce the dependency of the stock market on time. This would also aid in providing long term financial forecasts and improve the accuracy of the prediction.

Our project uses the method of Time series forecasting to predict the future price of the stock. The time attribute of stock data, given the uniqueness of it, is considered as an indispensable part of the prediction model. which is considered to be the main factor to be evaluated by utilising the open source library prophet. The dataset[2] that is used for the project contains stock prices of the famous banking institution, Santander Group from the Euronext Stock Exchange. It contains time series data for over a period of four years and also the attributes of the stock on each day such as highest value, lowest value, open value and the last value. The data set will be filtered into trend, seasonality and other miscellaneous attributes in the data preprocessing phase and divided into training dataset and test dataset. The data pipeline further uses the prophet model which is configured with the training set as the input. The next few segments of the paper contain a detailed description of the above process. The second section encompasses all the previous work that had been carried for solving the problem of stock market prediction. The third, fourth and fifth segments are methodology, result, conclusion and future work respectively.

2. Related Work

The need for the creation of more accurate and robust stock market prediction systems has motivated countless researchers to uptake this complex task. Bin Cao et al. [3] combined fuzzy rough theory and evolutionary neural networks to produce evolutionary fuzzy rough neural networks. This innovative design led to the creation of a system that had interpretability and prediction capability. This system outperformed long short-term memory network and the existing fuzzy rough neural networks. Lei Shi et al. [8] bridged the gap between the end users and the text based deep learning models. They did this by the creation of a system named DeepClue which constructed visualisations of the prediction done by their model. Their prediction model extracted predictive factors which enabled the analysis of hierarchies affecting those factors. Ballings, M., et al.

[1] evaluated various algorithms ranging from single classifier models such as support vector machines to ensemble methods such as random forest . They did so by employing area under the receiver operating characteristics curve (AUC) as a measure of performance. Data was gathered from numerous European companies which were publicly listed. Their findings indicated that random forest outperformed the remaining algorithms and the least accurate algorithm was logistic regression. Their finding however could be limited to only european markets. Xiaodong Li et al. [15] tackled the challenge of how news driven stock prediction systems that had been trained by utilising stocks with very little financial news could be improved. They proposed a novel method which employed sentimental transfer learning wherein knowledge accumulated from stocks rich in news is transferred to stocks with little coverage. The transfer of insights was done keeping in mind that the stocks were similar enough to be compared. They did so by ensuring that the historic time series of the target and source stocks were correlated and both the stocks were from the same sector. Peng, D. et al. [5] studied how the volatility of the stock market is affected by investor sentiment. The volatility decomposition theory of Pollet and Wilson was utilised to conduct a comparative analysis. Their approach to tackle this problem involved the use of big data in the construction of their model. Their examination of this enormous data revealed relevant insights on business rules and association models. Their efforts provided a fresh perspective for the analysis of market volatility. L. Owen et al. [7] Proposed an innovative solution for the purpose of stock market prediction. They combined the powers of convolutional neural networks (CNN), Multi-Layer Perceptron (MLP) and Long Short - Term Memory (LSTM) to produce SENN, which stands for Stock Ensemble-based Neural Network. Their model was then trained on the sentiment score extracted from stock based microblog and the stock data of Boeing. They found that the integration of sentiment score could greatly reduce the error percentage and increase the performance rate. M. Nabipour et al. [9] focussed on reducing the risk involved in stock market prediction by employing various deep learning and machine learning algorithms. Their study compared a total of nine algorithms such as Support Vector Classifier (SVC), Artificial Neural Network (ANN), eXtreme Gradient Boosting (XGBoost) etc. In order to conduct their analysis they chose four different stock market groups. They found that Recurrent Neural Network (RNN) and Long short-term memory (LSTM) performed considerably better than the other algorithms, they also found that deep learning methods performed significantly better. Y. Ji et al. [17] presented a hybrid model integrating long-short term memory (LSTM) and improved particle swarm optimization (IPSO) for accurate stock price prediction. They conducted their analysis on the Australian stock market (ASM) dataset, the ASM has a significantly higher return rate compared to its counterparts such as the NYSE. The combination of IPSO and LSTM proved to have performed better than other models such as the support-vector regression, PSO-LSTM and LSTM. D. Cabrera et al. [4] combined the domain of psychology and technology by integrating resilience and artificial autonomous systems. Resilience in psychology is defined as the ability of an individual to overcome any hurdles. This amalgamation resulted in an intelligent system that has the ability to overcome any problems it can encounter. Their ingenious approach proved to be fruitful as their system was able to withstand challenges and make appropriate decisions. M. Wen et al. [10] designed an algorithm which united pattern (motif) based sequence reconstruction and convolution neural networks (CNN) to combat the highly volatile nature of the stock market and its non linear relationship with time. They proposed that identifying patterns can help in the identification of a stock trend. Their proposal turned out to be beneficial as it outperformed existing sequential models such as LSTM which is basically a Recurrent Neural Network (RNN) in terms of computational complexity. X. Zhou et al. [16] modelled a novel method in order to gauge yield prediction. Their approach involved the application of virtual reality (VR) technology to construct a virtual stock trading scene. They designed their model based on long short-term memory, LSTM was chosen as it is good at processing time series based data. They designed the system in such a manner that an investment only occurs if it does not cause shorting in the virtual reality simulation. Their results proved that LSTM reigns superior in terms of time series prediction models, the same cannot be said for support vector machines. W. Cao et al. [14] highlighted the significance of considering the intricate relationships between the stock markets of different countries. Existing work done to solve the problem of stock market prediction often neglect to consider this important factor. In order to examine these relationships analysis has to be done on existing market dynamics. To solve this challenge they proposed a solution called Multi-layered Coupled Hidden Markov Model (MCHMM). This system was tasked with identifying the inner working of a particular country's market and the interplay of markets affecting that country. Their innovative approach performed significantly well in terms of business and technical perspectives. G. Liu et al. [6] emphasised the importance of analysing numerical data for accurate stock market prediction. They wrote about how numerical data carries more gravity as compared to news based models and in order to increase accuracy both these factors should be considered and are in fact complementary to each other. Their approach analyses the complementary relationship between news and numerical data. This novel idea filters out the noise and identifies trends from news related data performing better than certain baseline models. Sheeba SL et al. [11], [12] and [13] uses time indexes to tag in data.

3. Methodology

Harnessing the power of stock market prediction has innumerable benefits ranging from knowing what stocks to buy and sell in order to increase our profits at an individual level to helping in the formulation of economic policies at a government level. The process of reaping these benefits can be quite challenging as each step of this process is equally significant and should be carried out in an orderly manner. To reach the desired results certain measures have to be taken like analysis of the data, pre-processing and the final step prediction. All of these processes are explained in the following sections.

3.1. Understanding the Dataset

Understanding the dataset helps us garner a better understanding of what we are dealing with and how to best utilise it. The task of stock market prediction was carried out by utilising the Banco Santander, S.A. dataset.[2] It consists of the stock prices of Banco Santander whose currency is EUR and belongs to the Euronext Lisbon market. The dataset[2] contains 1,152 total number of rows where each row contains information regarding the date, the opening value, the highest value, the lowest value, volume and the turnover, This myriad of information spans through a period of four years starting from 2014/02/14 and ending on 2018/09/28.

Since the date attribute is given, we can even conduct analysis based on daily, weekly, monthly, quarterly and annual basis. The data in the given dataset[2] was separated by the "," delimiter. All the rows have values and there were no null or missing values. The data was completely numeric containing only float and integer values. We will be utilising the describe() function found in pandas in order to obtain various summary statistics of our data. These steps would help us garner valuable insights that can be utilised for improving accuracy of our model. We will be dealing with the closing stock prices of our dataset[2] and in order to gain insights on its relation with time fig. 1 was plotted, the SMA for a period of 5 days and 100 days was also plotted as depicted in fig.2.

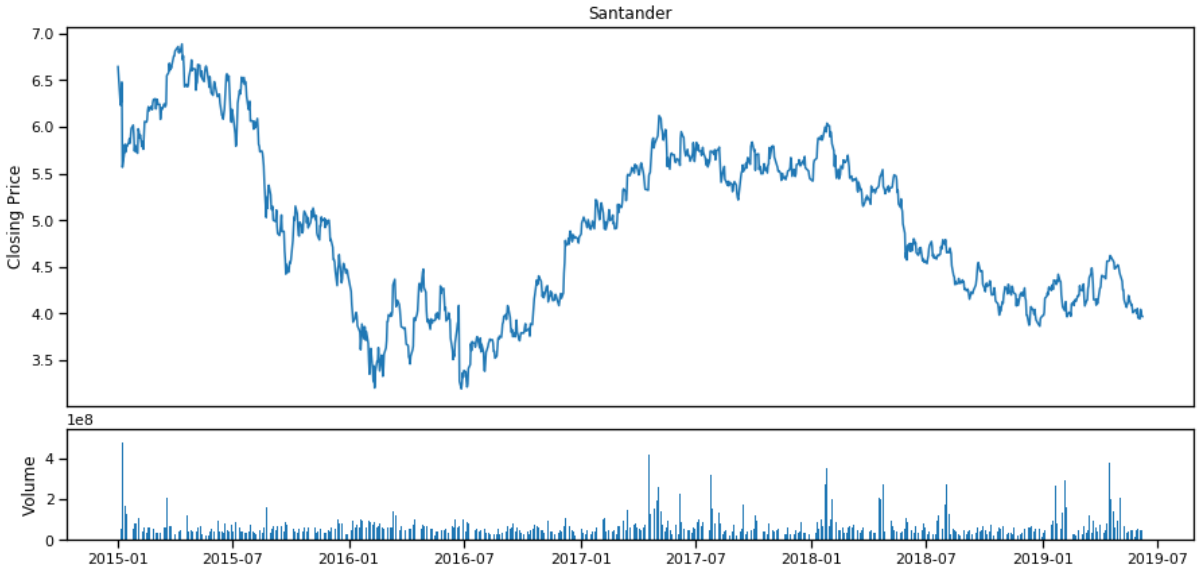

Figure 1. Closing Stock Prices

The simple moving average aids in reducing the noise associated with stock prices. It also helps in better understanding the impact played by old stock price values for the prediction of future prices. The blue line in fig.2 represents the closing stock prices of Banco Santander, S.A.

The orange line indicates the simple moving average which is taken for a time period of five days and the green line represents the simple moving average for a time period of 100 days. It is evident from the graph that SMA for a five day time period has a significant overlap with the actual values. This is in contrast to the SMA for 100 days which hardly had any overlapping values. This indicates that for prediction of the stock price its imperative to consider values from a shorter time period.

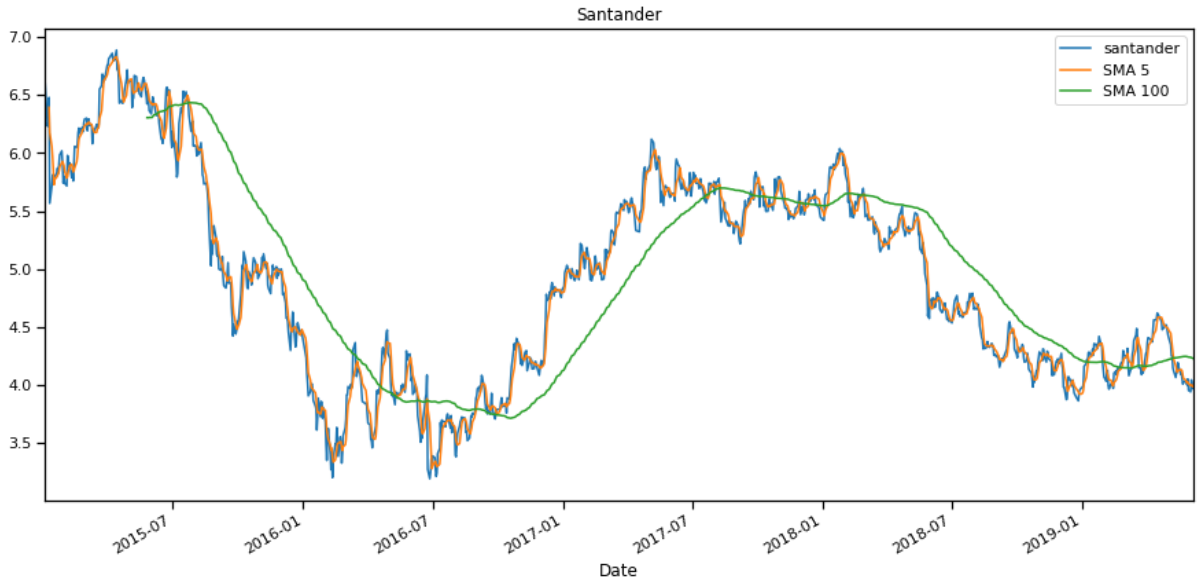

Figure 2. Simple Moving Average for 5 days and 100 days

3.2. Data Preprocessing

Data preprocessing is a crucial step which cannot be neglected. In this step we will be converting the data contained in our dataset[2] such a way that the machine can understand it, in other words the machine can easily parse it. The data would be split into test and training dataset.

3.3. Prophet

The time series data though extremely valuable and insightful requires the employment of a suitable algorithm . The prophet algorithm is one such algorithm which successfully extracts the required information. Prophet is a decomposable model meaning it breaks down a complex problem, such as prediction of time series data, into smaller problems. In order to do this, it considers three parameters namely seasonality, holidays and trend.

$$y(t) = g(t) + s(t) + h(t) + \varepsilon t$$

Here, g(t) stands for trend, s(t) stands for seasonality, h(t) stands for holidays, εt is the error rate. The trend parameter is tasked with tracking two more parameters namely saturation growth and changepoints. Saturation growth keeps track of how consistent the data is, for example the number of customers in a shop might reduce when its competitor opens a shop close by. The changepoints parameters keeps track of any sudden change that might cause a significant increase or decrease of stock price. A major difference between prophet and traditional approaches is the fact that it attempts to fit additive regression models, in short it tries to fit the curve. It can withstand data showing yearly, monthly, weekly and even daily seasonality, the plot for yearly seasonality is depicted in fig. 3. Seasonality is another parameter that's considered by prophet which utilises fourier series to provide an accurate end model.

$$s(t) = \sum_{n=1}^{N} \left(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right)$$

Here, s(t) stands for seasonality, P refers to the time period which can be taken at a monthly,weekly,daily, quarterly and even annual basis. N refers to the frequency of changes and the parameters a_n, b_n depend upon the given N. Identification of seasonality (repetitive patterns) is extremely crucial for extracting significant insights which would help in building much more accurate models. Events and holidays are also considered. During certain festivities changes may occur which would have an impact on the forecasts that are to be made. The final prediction is depicted in fig. 4.

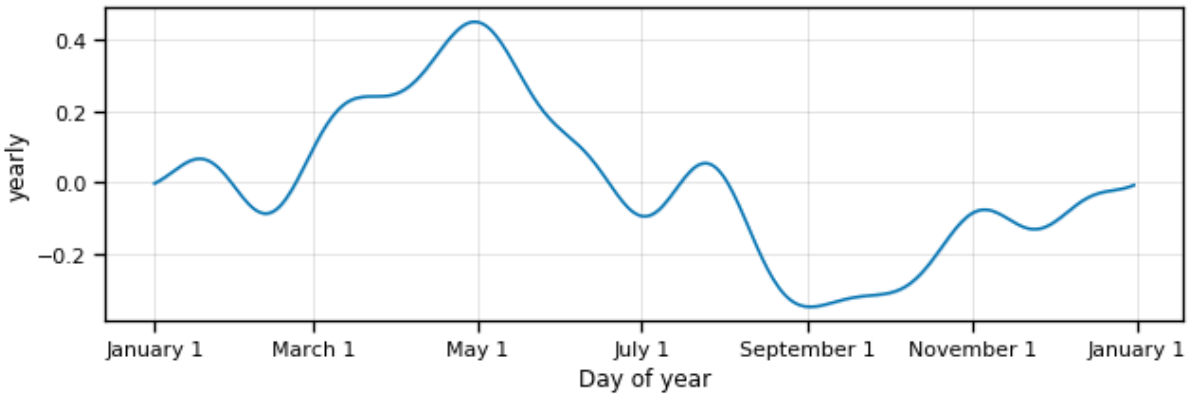

Figure 3. Yearly Seasonality

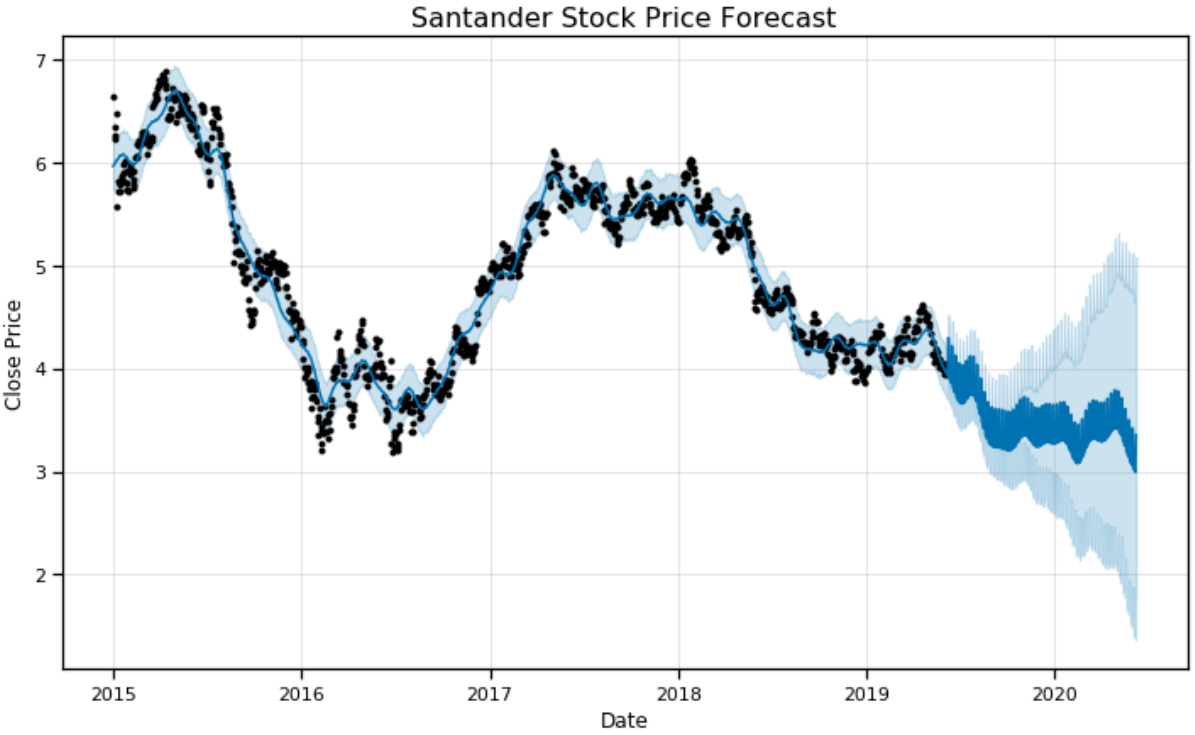

Figure 4. Stock Price Prediction

4. Result

Our work highlighted the role played by time in stock market prediction. Time series data was utilised for training and testing the model. The prophet library was employed to extract the useful information which yielded successful results. Its ability to handle seasonality proved to be extremely fruitful and helped in minimising the uncertainty associated with the prediction of stock market prices. Our study found data which was nearer to the current time had a significant impact on prediction meaning that as time passes on, the historic values of a stock seem to be less significant as the current price. It was also found that the stock prices for our dataset[2] increased significantly during the month of May and had a slight increase on Thursdays.

5. Conclusion and Future Work

Living in a constantly evolving world, it has become a necessity to adapt to the world as change has become the norm. Harnessing the power of predicting the volatile stock market can bring some stability in an individual's life. To do so requires the careful consideration of several factors, but the most important one of them all would be time. A novel method to exploit time of its insights is by employing the Prophet library. We concluded that historic data holds little significance in terms of prediction and the closing price increased on certain months and days of the week. The complexity of time and the variance of seasonality has to be studied in depth in order to improve the existing models. Our future work would focus on studying the other impactful factors affecting the stock market and how best they can be integrated with time.

References

[1]. Ballings, M, Dirk Van den Poel and Nathalie Hespeels, Ruben Gryp (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications, Elsevier Ltd.*
[2]. Banco Santander dataset from Quandl website, https://www.quandl.com/data/EURONEXT/SANT
[3]. B. Cao, J. Zhao, Z. Lv, Y. Gu, P. Yang and S. K. Halgamuge (2020), Multiobjective Evolution of Fuzzy Rough Neural Network via Distributed Parallelism for Stock Prediction, *IEEE Transactions on Fuzzy Systems, Volume 28.*
[4]. D. Cabrera, R. Rubilar and C. Cubillos (2019). Resilience in the Decision-Making of an Artificial Autonomous System on the Stock Market, *IEEE Access, Volume 7*.
[5]. D Peng (2019). Analysis of Investor Sentiment and Stock Market Volatility Trend Based on Big Data Strategy, *International Conference on Robots & Intelligent System (ICRIS), Haikou, China.*
[6]. G. Liu and X. Wang (2018). A Numerical-Based Attention Method for Stock Market Prediction With Dual Information, *IEEE Access, Volume 7.*
[7]. L. Owen and F. Oktariani (2020). SENN: Stock Ensemble-based Neural Network for Stock Market Prediction using Historical Stock Data and Sentiment Analysis, *2020 International Conference on Data Science and Its Applications (ICoDSA).*
[8]. L. Shi, Z. Teng, L. Wang, Y. Zhang and A. Binder (2018). DeepClue: Visual Interpretation of Text-Based Deep Stock Prediction, *IEEE Transactions on Knowledge and Data Engineering, Volume 31*
[9]. M. Nabipour, P. Nayyeri, H. Jabani, S. S. and A. Mosavi (2020). Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis, *IEEE Access, Volume 8.*
[10]. M. Wen, P. Li, L. Zhang and Y. Chen (2019). Stock Market Trend Prediction Using High-Order Information of Time Series,*IEEE Access, Volume 7.*
[11]. Sheeba SL and Yogesh P (2020). Enhanced Cache Sharing through Cooperative Data Cache Approach in MANET , *International Journal of Biomedical Engineering and Technology, Inderscience, Volume. 32.*
[12]. Sheeba SL and Yogesh P (2015). A Novel Context Aware Counter Based Cooperative Cache Replacement Strategy for Mobile Networks, *International Journal of Applied Engineering Research, Volume. 10.*
[13]. Sheeba SL and Yogesh P (2020). A Time Index Based Approach for Cache Sharing in Mobile Adhoc Networks , *Proceedings of the international conference on computer science, engineering and applications, Volume 1*.
[14]. W. Cao, W. Zhu and Y. Demazeau (2019). Multi-Layer Coupled Hidden Markov Model for Cross-Market Behavior Analysis and Trend Forecasting, *IEEE Access, Volume 7.*
[15]. X. Li, H. Xie, R. Y. K. Lau, T. Wong and F. Wang (2018). Stock Prediction via Sentimental Transfer Learning, *IEEE Access, Volume 6.*
[16]. X. Zhou, M. M. Kamruzzaman and Y. Luo (2020). Mathematical Model of Yield Forecast Based on Long and Short-Term Memory Image Neural Network , *IEEE Access.*
[17]. Y. Ji, A. W. -C. Liew and L. Yang (2019). A Novel Improved Particle Swarm Optimization With Long-Short Term Memory Hybrid Model for Stock Indices Forecast (2021), *IEEE Access, Volume 9.*