International Journal of Advance Research in Science and Engineering
Volume No.06, Special Issue No.(01), Nov 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

Forecasting Time series Market Data Using Machine Learning: A Literature review

Asadullah Khan[1], Dr. Khalid Mohiuddin[2]

[1] *PhD Scholar Faculty of Computing & Information Technology, Himalayan University,(India)*

[2] *Department of Information Technology, College of Computer Science, KKU, KSA.*

ABSTRACT

In this paper, previous studies featuring, Machine learning based stock market analysis and predictions have been reviewed. This study is done to examine the various methodologies used in analyzing stock market data, and methods used prediction and forecasting the market. We propose a methodology that can be used in forecasting time series market data. Also we study the data sources. We have examined several publications and many theses in this area. Selected main features which are applicable for further studies are mentioned. Presently, Social media including Twitter, Facebook, Web contents, has provided enormous unstructured opinion contents for various decision making process. Sentimental analysis is also necessary for market predictions. Efforts are put to study the latest methods used in sentimental analysis of web contents and social media.

Keywords: *Forecasting Stock market, Machine learning, Market prediction, review, Time series forecast.*

I. INTRODUCTION

Forecasting stock market price is a very attractive and challenging task among business, economists and academia. There are two major sources for analysis. One is historical market data and the other market sentiments. The latter is unformatted information such as news about company, politics and any other social factor, etc. Research is being done to evaluate the sentiments, opinions, based on the huge information available on Blogs, Twitter, analysis reports etc. Studies done on use of data mining algorithms and use of machine learning to understand the behavior of market and to predict the market , But not much of work has been done to study the Indian stock markets and to predict the near future market price. Many even believe the stock markets are not predictable.

The stock market responds to many factors. It is possible to watch at the stock at various instances when it behaves interestingly, this data can be presented as a time series. Then analysis of the time series is done and an appropriate model can be used to fore cast the prices which possibly help the investor make decision when to sell or buy. Machine learning techniques can be used as model to forecast. Machine learning a sub-field of computer science is the study and application of computers that possess the ability to find patterns, generalize and learn without being explicitly programmed. In this paper we review the work done on these aspects.

This paper contains 4 sections, section2 is review of literature and section 3 lists various algorithms applicable and section4 is our methodology and conclusions.

II. REVIEW OF LITERATURE

Jigar Patel et al. in their paper focuses on the task of predicting future values of stock market index. Two indices namely CNX Nifty and S&P Bombay Stock Exchange (BSE) Sensex from Indian stock markets are selected for experimental evaluation. Experiments are based on 10 years of historical data of these two indices. The predictions are made for 1–10, 15 and 30 days in advance. The paper proposes two stage fusion approach involving Support Vector Regression (SVR) in the first stage. The second stage of the fusion approach uses Artificial Neural Network (ANN). [1]& [2]. They have taken technical features which derives the market, However other factors which equally affects the volatility of market is not considered such as the market sentiments and news which is present in abundant in social media.

Arman Khadjeh Nassirtoussi et al. Review the related works that are about market prediction based on online-text-mining and produce a picture of the generic components that they all have. And compare each system with the rest and identify their main differentiating factors.[3]

Xue Zhang et al. used text content from twitter to predict stock market indicators for Dow Jones, NASDAQ and S&P that is U.S. stock market. [4]. such a study has not been done on Indian stock market.

Weiling Chen et al. used news from Chinese largest online social website Sina Weibo to predict Chinese stock market. They have implement a model based on Recurrent Neural Networks (RNN) with Gated Recurrent Units (GRU) to predict the stock volatility in the Chinese stock market. [5]. Such a study has not been done on Indian stock market.

Shoiab Ahmed and Ajit Danti proposed to produce the word counts and categorizing them into seven classes as strong-positive, positive, weak-positive, neutral, weak-negative, negative and strong-negative words using SentiWordNet. This is used to analyze the sentiments from web contents [6].

Michel Ballings et.al. Compares and benchmark ensemble methods (Random Forest, AdaBoost and Kerrnel Filter) against single classifier models (Neural Networks, Logistic Regression, Support Vector Machines and K-Nearest Neighbor). The main contribution of this study is an extensive benchmark comparing the performance of ensemble methods (RF, AB and KF) and single classifier models (Neural Networks (NN), Logistic Regression (LR), SVM, K-Nearest Neighbors (KNN)) in predicting the stock price direction. They shows that, given their superiority in other domains, ensemble methods will outperform the single classifier methods [7]

Osman Hegazy et.al, In this Work they have used historical data and technical parameters as input data and use ANNs to predict the results.[8].

Krzysztof Jerzy Geras in his mater thesis discussed various classifier and booster algorithms including Bayesian, AdaBoost, and Random Forest. He has shown the market prices are event driven. He writes, " When information is widely dispersed among economic actors, a prediction market offers a mechanism to collect and aggregate that information. Much of the enthusiasm for prediction markets comes from the efficient-market

hypothesis In a truly efficient prediction market, the equilibrium price is the best predictor of how likely an event is, and no other use of available information may lead to improved market-generated forecasts".[9].

Johan Agneman has analyse how the outbreak of European debt crisis has effected the volatility in Eurozone's stock in the region. It shows the market interdependencies among markets. [10]

III. ALGORITOMS

In literature we have many Data mining algorithms and still they are applicable for internet of thing and hence can be used in machine learning and its applications. One of the important aspects in stock prediction is to classify the data. The experiment results of 8 data mining algorithms and their performance in classification accuracy CA is shown in table1.

Support Vector Machine (SVM),

K-nearest neighbor (KNN),

Nave Bayes (NB),

Linear discriminant analysis (LDA),

C4.5,

C5.0 ,

ANNs (Artificial Neural Networks),

and Deep Learning ANNs (DLANNs) .

They conclude the best Classification accuracy CA is achieved in DLANNs. Fine tuning the parameters in DLANNs enhances CA. DLANNs use huge share of resources as they have a very complex structure.[10]

Table1. Classification accuracy in %

Datase1		Datase2		Datase3	
Algorithm	Accuracy %	Algorithm	Accuracy %	Algorithm	Accuracy %
SVM	98.5	2350.1	86.43	5.2	91.75
KNN NB	98.94	450.6	86.88	0.88	78.67
C4.5	77.04	0.52	70.09	0.02	52.72
C5.0	99.69	22.65	91.81	0.15	99.95
LDA	99.62	21.1	90.26	0.13	99.96
ANN	81.85	0.98	71.53	0.02	66.4
DLANN	99.03	33228.1	89.55	94.2	100
	99.52	12600	87.10	210.12	98.49

Study shows that many existing methods proposed for the task of forecasting market price are direct applications or modifications of one or more standard machine learning methods for classification, dimensionality reduction, density estimation and regression.[11]

Study shows that many existing methods proposed for the task of forecasting market price are direct applications or modifications of one or more standard machine learning methods for classification, dimensionality reduction, density estimation and regression. The study show computational complexity of learning and prediction Methods are shown in the following Table.

Table Computational Complexity Of Learning and Prediction Methods

Method	Learning	Prediction
PKNN	$O(1)$	$O(NM + N \log N + KM)$
NBClass	$O(NM^2V^2 + M^2 \log M)$	$O(MVK)$
K-Medians	$O(INMK)$	$O(MK)$
wSVD	$O(INM^2 + IM^3)$	$O(IKM)$
Multi	$O(NMV)$	$O(1)$
MixMulti	$O(INMVK)$	$O(MVK)$
Aspect	$O(INMVK)$	$O(MVK)$
URP	$O(I_1I_2NMVK)$	$O(I_1MVK)$
Attitude	$O(I_4I_2I_1NMVK + I_4I_3NMVK)$	$O(I_2I_1MVK + SMVK)$
N: #users M: #items V: #rating values I: #iterations S: #samples K: complexity		

The lowest computational complexity is neighborhood method PKNN-Learn, The complexity of PKNN-predict method scales with number of users N rises.

Fundamental tradeoff in computational complexity between the learning algorithm and the prediction algorithm. From a systems stand point it is much more desirable to tradeoff higher learning time for lower prediction time. Learning can be done offline, while prediction must be done online, and often in real time for Internet-based recommendation services.[11]

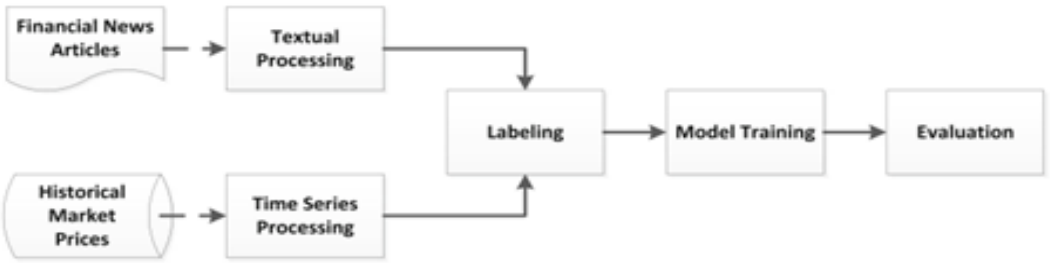
IV. PROPOSED METHODOLOGY AND CONCLUSION

Our study shows that the existing machine learning algorithms and machine learning techniques can be used in forecasting time series market data. We will outline our methodology for further work for this task.

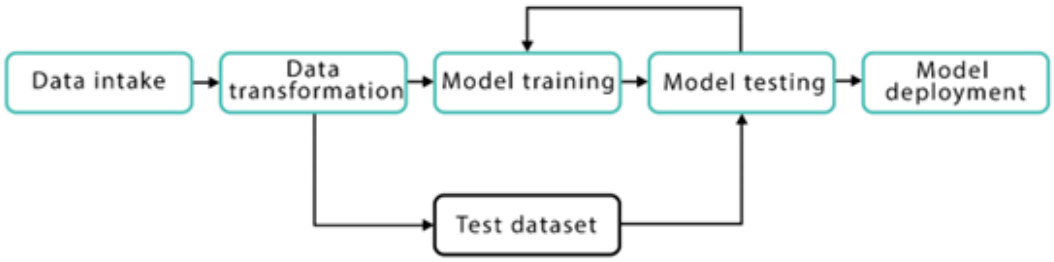
We will Use Historical Data of stock market, Technical /Financial parameters Of Company, News from various sources. NLP is used to generate structured information from news Source. Machine learning Techniques are applied to Structured Data to produce the desired results.

Architecture of traditional approaches is depicted as follows

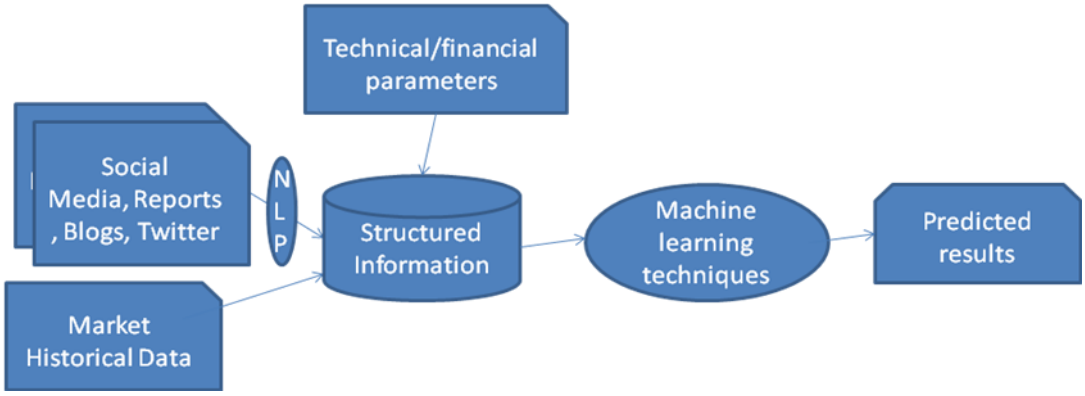
International Journal of Advance Research in Science and Engineering
Volume No.06, Special Issue No.(01), Nov 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354



General workflow of machine learning process is depicted as follows.



Proposed Model is depicted as



REFERENCES

[1] Jigar Patel, Sahil Shah, Priyank Thakkar , K Kotecha "Predicting stock market index using fusion of machine learning techniques" Expert Systems with Applications 42 (2015) 2162–2172

[2] Jigar Patel, Sahil Shah, Priyank Thakkar , K Kotecha "Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques" Expert Systems with Applications 42 (2015) 259–268

[3] Arman Khadjeh Nassirtoussi , Saeed Aghabozorgi , Teh Ying Wah , David Chek Ling Ngo "Text mining for market prediction: A systematic review" Expert Systems with Applications 41 (2014) 7653–7670

[4] Xue Zhang, Hauke Fuehres , Peter A. Gloor "Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear" COINs2010: Collaborative Innovation Networks Conference

[5] Weiling Chen, Yan Zhang, Chai Kiat Yeo, Chiew Tong Lau, Bu Sung Lee "Stock market prediction using neural network through news on online social networks" Smart cities Conference(ISC2) 2017 International IEEE *Xplore*: 02 November 2017

[6] Shoiab Ahmed and Ajit Danti "A Novel Approach for Sentimental Analysis and Opinion Mining based on SentiWordNet using Web Data" 2015 International Conference on Trends in Automation, Communications and Computing Technology (I-TACT-15)

[7] Michel Ballings, Dirk Van den Poel, Nathalie Hespeels, Ruben Gryp "Evaluating multiple classifiers for stock price direction prediction" Science Direct Expert System with applications 15 November 2015.

[8] Osman Hegazy , Omar S. Soliman and Mustafa Abdul Salam " A Machine Learning Model for Stock Market Prediction" International Journal of Computer Science and Telecommunications [Volume 4, Issue 12, December 2013].

[9] Krzysztof Jerzy Geras "Prediction Markets for Machine" Learning Master thesis in COMPUTER SCIENCE, September 2011

[10] Johan Agneman "Stock market interdependencies" Master's Thesis in Financial Economics University of Gothenburg School of Business, Economics and Law. 2013

[11] Furqan Alam, Rashid Mehmood, Iyad Katib, Aiiad Albeshri "Analysis of Eight Data Mining Algorithms for Smarter Internet of Things (IoT)" ScienceDirect , Procedia Computer Science 98(2016)437-442

[12] Benjamin Marlin "Collabora tive Filtering: A Machine Learning Perspective" A thesis submitted in conformity with the requirements for the degree of Master of Science Graduate Department of Computer Science University of Toronto.