Pakistan stock exchange prediction using RIDOR classifier

Wasim Akram, Muhammad Imran *

Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad, Pakistan

ABSTRACT

Now-a-days stock market prediction is important activity and interesting topic for professional analyst. The stock market is biggest investment platform, however investment in stock market need accurate and complete information. The accurate and fast-prediction of stock market attracted the investor for profitable output. The stock market prediction is complex task because uncertainty involves in the market movement up/down. Mostly the machine learning techniques (MLT) are used for accurate prediction of stock market, because of its capability of partitioning; extract hidden information form raw data, monitors the fluctuation rate of stock market, suitable for nonlinear data etc. This research work is about to review the strength and weakness of existing stock market prediction techniques. This research work proposed a Ripple-down-rule-learner (RIDOR) classifier based technique. The RIDOR rule base classifier generates default value and work like if-else statement for uncertainties. The other contribution is a prepared data-set using technical indicator to predict stock market trend. The output of the propose model is outperformed as compare to the exiting techniques.

© 2017 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

The process of making decision about the future performance by using monthly, daily, hourly or historical data is called prediction. Now a day prediction/forecasting are important and interesting research area. Prediction used (Babu and Reddy, 2014) in different application such as internet-traffic facilities service providers for enhancing their service, prediction temperature, weather and change in environment-gives facilities to formals or agriculture sectors, prediction disasters such as earth-quick, flood etc. The prediction is important activity in stock market which gives information to investor for safely investment in the stock market. The stock-market involves different data such as seasonal and non-seasonal, highly-volatile and less-volatile, linear and non-linear, Gaussian and non-Gaussians data. Prediction provided- information about the stock-market, which used investors to build favorable scenarios in futures such as return of investment, risk of investment, and also used to build-decision making strategy. The stock market prediction is complex task because uncertainties involve in the stock market data, which future decrease the performance in short term, medium term and long term. Before investing the stock market the investor performed two different types of analysis (de Oliveira at al., 2013).

- Fundamental analysis: the investor looking the basic value of stock-market, performance of companies and its economy, politics involvement-scenario etc.
- Technical analysis: in technical-analysis the investor calculate the stock for the purpose of studying market-activity such as prices and volumes.

The proper selection of prediction methodology has been important for proper planning and issues which may arise in businesses institution. Financial-stability of businesses institution is mostly depending on accurate prediction, which is further used for making the key decision about stock market. Mostly stock market prediction relays only on accuracy and don’t care about the prediction speed. However stock-market prediction speed is important factor for stock market, because a mini second affected stock market. The prediction process complexity depends upon the various input variables which consistent various factors. The historically, monthly, daily and hourly-patterns of variable have different properties which is used as input to prediction process but actually it is more complex.
There are many factors which affect the stock market performance. The most important factors include: inflation rates, interest-rates, employment-rates, the real-estate-market, oil-prices, war, natural-disasters, big-company-mergers, big company outs buy and good/bad company news (Patel et al., 2015a; Asadi et al., 2012).

In this research work proposed a RIDOR technique for market prediction. The stock market prediction involves with uncertainty which rule base classifier manage accurately (Qin et al., 2009). Different technical indicators are used as input to prediction models. The rest of this paper include the literature review, proposed approach, data set, data preparation, experiments and results, and conclusion and future work.

2. Literature review

In last few-years the stock-market is interesting field of research. Many more works are proposed by researcher to predict the stock. In this section include the literature of different research article which publish in well-known publication journal are given below.

Kara et al. (2011) explored SVM and ANN for prediction of prices-movement in stock-market. Three layered feed-forward-NN are built for input and output. The SVM merge with learning-theory for optimal trade-off in b structure-complexity as well as risk. The SVM hyper-plane distributed positive and negative value. The output 0-shows down-direction while 1-shows up direction of stock market. The accuracy of proposed techniques in the term of point-of-change-in-direction (POCID) of both NN and SVM was 75.74% and 71.52% respectively. There are no rules for parameters setting for sensitive-parameters and comprehensive-parameter for both models.

Patel et al. (2015a) used ANN, SVM, RF, and Naïve-Bayes for accurate prediction. In this research the author introduce Two-approaches for input these models. In first approach ten-technical-indicators were calculated by (open-price, low-price, and high-price, close-price) which use as input to these models. In the second approach trend-deterministic data-preparation-layer was used to converted continuous-data to discrete one by using the input indicators which show up/down of the stock market. To validate this work the author used historical-data and compares the result of these models. It is found that the proposed input-approach show-best-result in the term of accuracy. Its accuracy in ANN was 86.69%, SVM was 89.33%, RF was 89.98% and Naïve-Bayes was 90.19% respectively. The trend-deterministic-layer preparation-layer both risk and return are categories into two different parts which are complicated in measurements.

Patel et al. (2015b) proposed two stage prediction approaches for prediction-future stock-market index. In This research works two different layer was used; in first layer support-vector-regression (SVR) was used to predict future-value. The future value was used as input in 2nd layer and combines it with the ANN, Random Forest and SVM for the prediction models. The result of proposed-approach was compare with single stage approach. It is found that the proposed-approach show batter-performance in the term of accuracy. It accuracy improved ANN vs. SVR-ANN was 11.5%, SVR vs. SVR-SVR was 1.66% and RF vs. SVR-RF was 9.1% respectively. The proposed model is accurate and robustness, may reduce performance whenever increase parameters.

Asadi et al. (2012) proposed-hybrid intelligent model called pre-processing evolutionary Levenberg Marquardt-neural-network (PELMNN) for stock-market-prediction. Frist Genetic-algorithm (GA) was used to find optimal-weight for artificial-neural-network (ANN). The pre-processing technique was used twice once to minimize input of Levenberg Marquardt-BP, and second used for transformation of data. To validate PELMNN the author used 50 stock-indexes. The result of proposed techniques was compared with hybrid fuzzy model and ANN, the found that the proposed PELM-NN achieve batter performance in the term of prediction accuracy in Mean.Absolute-percentage-error (MAPE) was 0.51% where the MAPE of hybrid fuzzy model and ANN was 1.3% and 0.78% respectively. The evolutionary algorithms slow down speed whenever used in other techniques.

Qiu et al. (2016) proposed a hybrid-approach to predict Japanese-stock-market. In this approach new data-set was introduce as input to maps non-linear data. In this approach classical-back-propagation-learning-algorithm was used for efficient return. In this approach GA and simulated annealing (SA) was used to get-better in prediction while preprocessing used to reduce search space and fuzzy-curve analysis used to find-out correlation in input-output variable. To validate this approach the author used historical data. The result of proposed-model was compare with BP-NN. It is found that proposed model show batter result in the term of accuracy and minimize errors. Its accuracy in best case was 0.0725 by using 28 CPU times where BB-NN was 0.0044 by using 68 CPU time. Whenever variable gives as input the Lag (delay) issue arises.

de Oliveira at al. (2013) presented ANN base model for predicting financial market. This model is the combination of economic and financial-theories, which used technical and fundamental-analysis of time-series.

These technical and fundamental-analyses was used to predict the future behavior of stock-price. Four different functions performed such as Understood problem-domains which point out the key variable, pre-selection and getting-samples, preprocessing are used to as input, and finally design a model for predict stock market. The proposed-techniques show best result Point-of-change-in-direction POCID of 93.63% with window size of 3, and MAPE is 5.95%. The proposed work calculated the behavior and trend of price and eliminate to identify risk and return.
Chang (2012) proposed partially-connected-NN architecture with some new features. The NN connected with different neurons, its weight was selected by using some probabilistic-data, and used more hidden layer store more information. In this architecture evolutionary-algorithm was used to improve the learning and training-weight while The GA was used for global-search and train weight. To validate the proposed-architecture the author used probabilistic-data which achieve better performance in the term of accuracy. Its accuracy in training was 94% and testing was 97% respectively. The data of time series is directly inputted to the proposed model, used some suitable techniques such as preprocessing for properly input the time data.

Wang et al. (2012) introduced a new hybrid approach for prediction price-index. The proposed approach used-exponential-smoothing method (ESM), BPNN, auto-regressive-integrated-average (ARIMS) and GA which generated-weight for prediction model. This approach was captures-linear and non-linear properties for future-prediction. The ARIMS was used to predict the time series while ESM was used for find-out the change rate during fluctuation of stock market. For validate the author-used closing and open price of two different indexes average time was used. The proposed approach shows better-result in the term of accuracy. Its accuracy in ESM (MAE was 3991.5), The ARIMA accuracy (MAE was3615.8) and BPNN-model accuracy (MAE was 4453.6). The proposed work is time consuming and risky process when optimizing the parameters.

Ticknor (2013) proposed a Bayesian-reguarization-ANN to predict the behavior of stock market. Three layers feed-forward-NN are used for input and output. The Bayesian regularization have many properties such as reduce the issue of over fitting and over training for noise data, used for probabilistic weight to network, provided optimal result in a complex-system, improve the prediction quality etc. The author used stock indexes such as Microsoft and Gold-Sachs. The observed result of proposed technique achieved better performance in the term of accuracy. The Microsoft index training, testing and total of 1.0494%, 1.0561%, and 1.0507% respectively, while Gold-Sachs index training, testing and total is 1.5235%, 1.3291%, and 1.4860% respectively. This work focuses accuracy and avoids the efficiency which important for critical environment.

Choudhury et al. (2014) proposed new self-organization-map (SOM) hybrid clustering (K-mean-clustering) technique for stock market return. The K-mean-clustering and SVM was use for selection of portfolio. The best portfolio generate when accurate prediction of price and its variation are predicted. The SOM built neighborhood-relationship among the neuron, the k-mean cluster was used for clustering. The SVM was used to find-future value, and then prepared trading strategy. The proposed techniques was applied into different stock in the term of stock-return. The proposed techniques achieve better performance in SVM and ANN. Its return on investment was 19.6%, 8.2% and 18.2% respectively in all stock when using SVM and NN performance was 3.4%, 7.4% and 8.1% respectively. The SVM worse performed when increase data-set.

Babu and Reddy (2014) proposed auto-regressive integrated moving average-ANN to predict stock-market price. The ARIMA-ANN model works on the bases of one-step-ahead and multi-step-ahead prediction. The moving-average-filter (MAF) was used for decomposition of time-series data, MAF show the nature of data which applying ARIMA and ANN. The result compare with ARIMA and ANN model in one-step-ahead and three-step-ahead. The Proposed ARIMA-ANN model show best result in step ahead (MAE is 0.1884 and MSE is 0.0507) and multi-step a-head (MAE is 0.2951 and MSE is 0.1445). The MAF algorithm easy to use but reduces performance when using for noisy data.

Li et al. (2016) built a new trading-mining platform for prediction of stock-market. For speed and accuracy the author used back extreme learning-machine (BELM) and Kernelized extreme learning-machine (KELM) for collection hidden information in raw data which predict price-movements. The pre-processing techniques used as input for news articles and price. The normalization process converted data in the form of (-1, 1) which is show up/down of price. The author used historical data, then result compare by BPNN, SVM, and BELM. The result of proposed technique was better performed in term of accuracy and time. Its accuracy in precision was 0.148 where BPNN was 0.165; SVM was 0.261 and BELM 0.149. More memory and more CPU resource are required for proposed techniques.

Dai et al. (2012) proposed a non-linear independent component analysis by using NN to predict Asian-stock market. The original predicting variables are given as input then divided the input variable into different levels, and then non-linear data are selected to input the BPN for prediction model. The NLICA shift input to different feature which identify original data which used ANN for prediction. To validate the author used non-linear data. The proposed method-achieved better performance in the terms of accuracy by used different statistical parameters. Its accuracy in RMSE was 50.44%, MAD was 39.78%, MAPE was 0.242% and RMSPE was 0.302%.The propose prediction model was the capability to extract the feature into independent source from simply observed non-linear mixture data, there is no relevant data mixing methodology are available.

Laboissiere et al. (2015) proposed a methodology for the prediction of max-min stock price of Brazilian power distribution companies. In this technique multilayer-perception-architecture merges with ANN, while Levenberg Marquardt was used for estimating Max-Min day price. The preprocessing technique was used to identify weight for business-days. Weight-moving-average (WMA) was used for analysis of time series, while American-dollar used as possible attributes to predict max-man-day price,
open-close price and best-bad price. In this mythology the author used historical-data. It is found that the proposed methodology achieve better performance in the term of accuracy. The accuracy of Man-day-price was 0.9% while Max-day-price was 2.1%. This model check the correlation of input attributes but Maybe possible that one-to-many relationship between input variables.

Karimi et al. (2014) used ANN and GA for prediction of Tehran-stock-index. The objective of this research is to predict the stock-index for profitable out-come. In this study the author used ANN for predicting while the GA was used for optimizing the input-variable in the NN. To validate the proposed works the author used daily data from September-2010 to March 2013. The proposed work achieves batter result in the term of accuracy and reduces error whenever used 8-neuron in hidden layer. The accuracy increase and error reduce up-to 5% or less than 5%.

3. Proposed approach

In this research work the rule base classification algorithms are used to predict the future stock market trend. The rule base classifiers are mostly used for classification of Iris data set (Devasena et al., 2011; Veeralakshmi and Ramyachitra, 2015). The rule base classifiers also use for the prediction of biological and heart disease (Koklu et al., 2015; Lakshmi, 2014; Farid et al., 2016).

Now rule base classifiers used for the prediction of stock market trend and price moment of stock market (Shriwas and Sharma, 2014, Shriwas and Farzana, 2014). The stock market prediction depend on accuracy which we get form the correct classification of each data set. Mostly the stock market deal with uncertainties whenever the uncertainties occurs the stock market prediction failed, because the proposed method generated/classified the result in previous rule or previous training. The rule base classifier gives efficient result in uncertainties (Qin et al., 2009). The RIDOR (Veeralakshmi and Ramyachitra, 2015, Lakshmi, 2014) is rule base classifier which generated default rule for uncertainties whenever occurs during the prediction with minimum errors.

3.1. RIDOR

Ripple down-rule-learner (RIDOR) is a rule base classification which used for classification of Iris data set (Veeralakshmi and Ramyachitra, 2015) and heart disease prediction (Lakshmi, 2014). RIDOR classifier first generated the default-value for certain situation. Whenever the situation occur the default value generated with minimum error rate. The default value continuously iterated the value by using incremental rule which minimize the errors pruning and generated most accurate result. The RIDOR is also work on the concept of learning in the previous value and generated future value. The RIDOR mainly use the concept of if-else statement. RIDOR iterate value until its trues then generate output else generate default value as output. The RIDOR classifier is also useful for uncertainties because uncertainties values are set before iterating the training and testing data. The algorithm pseudo code for RIDOR classifier is given in Fig. 1. Fig. 2 shows the proposed prediction model.

Algorithm RIDOR (D, Rt)
Input: A relational database D with target relation Rt that contains P Positive and N negative tuple
Output: A set of rules for predicting class labels of target tuples.
Procedure:
Rule set R=empty
If [Rt] < MIN_SUP then return Ruler - empty rule
Set R, active
Repeat:
Find a rule in active relation branch and it not branch to active
R+r
relations into inactive
Learn except Set relation of r
R=X=X-r
Until (X=NUL)
Set all active
Return R
End

Fig. 1: Algorithm of RIDOR classifier

4. Data set

In this study obtained the data-set of Pakistan-Stock-Exchange from http://noormaier.net and https://finance.yahoo.com/website. The PSE data-set consist of (open, high, low, close and volume) from 1997-2016. The trading days from 1997 to 5 Dec 2016 is about 4656. The 20% of that data are used for selection parameter which exam to design parameters of prediction-model. The 30% of data is sub-divided into two different set one for training and one validation which show increase and
decrease of selection-parameters. The parameter-selection gives ideal parameters of prediction models. The whole-data is also sub-divided into two parts one for training and one of testing. Different types of technical indicators are used in this study which used as an input for prediction model.

Calculated each technical indicators in excel which show up/down according to the previous observation. We observed each technical indicator in three different perspectives in a time period of 3 days, 5 days and 10 days.

5. Data preparations

5.1. Simple moving-average

The simple-moving-average of stock-market is analysis tool which calculated the average of past days. In this study we exam the average of 3days, 5days and 10 days respectively. The calculation of all these simple moving average the current value compare to previous value if the current value is greater than previous value the stock market show down direction else up. Than we compare all the resultant value with each other if two of value show up then the stock-market goes up else show down-direction

=if (current-value>previous-value, show up, else show down)

5.2. Weighted moving-average

The WMA is same like the SMA which used for predicting short-term future value. We find-out the WMA of 3day, 5days and 10days. The current WMA compared with the previous value of WMA if it is greater than previous show up direction else show down-movement of stock-market. The resultant of all these three observation compared with each other if two of them show up direction than show up else show down-movement of stock-market.

=if (current-value>previous-value, show up, else show down)

5.3. Stochastic K% D% and Williams R

The STCK% STCD% and Williams-R% are stochastic-oscillators. We find-out 3days, 5day and 10 days of trend value all these indicators. When oscillators show increasing then the stock show-up (1) otherwise show-down (0), then we compare these 3days, 5days, 10days result of each stochastic-oscillators. When two value show up direction then show up else shows down-movement of stock market.

=if (current-value>previous-value, show up, else show-down)

5.4. Momentum

The momentum indicator is used for showing the fluctuation of stock market. In this study we find-out 3days, 5days and 10days of momentum which show fluctuation in the stock market into different scenario. If the momentum indicator show negative value than show down else show up movement of stock market. Than compared all the resultant value with each other if two of them show up its show up movement else show down-movement.

=if (current-value<0, show down, else show up)

5.5. Relative strength index (RSI)

RSI mainly used for the identification of overbought-oversold point. The resulting value of
3 days, 5 days and 10 days are higher than 70 show always down (0) direction, while the resulting value less than 30 show always up (1) direction. The value between 70-30 show up/down by comparing the current value. Finally the resulting value compare with each other if two of them show up show up-movement else show down-movements.

'=if (current-value>70, show-down, if (current-value<30, show-up, if (current-value>previous value, show-up, else show down)))

5.6. Price momentum oscillators

The decision point Price Momentum Oscillator (PMO) is an oscillator base on a Rate of Change (ROC) calculation that is smoothed twice with exponential moving average that used a custom smoothing price. Because the PMO is normalized, it can also be used as a relative strength tool.

'=if (current-value > 0 show up direction, else shows down direction)

5.7. Moving average convergence divergence (MACD)

The MACD also used for stock market, when its goes up direction the trend-direction of stock is goes up (1) else the stock trend is goes down and indicated (0). We also the used the same procedure as used in previous section.

'=if (current-value>previous -value, show-up, else show-down)

5.8. Commodity channel index

The CCI mainly calculated the difference of stock-price and its change with respect to average-price-change. The changes in change which respect to average show the strength and weakness. In this research work we divided by 1000 in total change in price. Then we selected the overbought value is greater than -200 and oversold value is 200. In between-value among (200,-200) is depend on the increase in decrease according to the comparison of current value to the previous value. We find-out the different three values of different days. We compared these value if two of them show up-direction the resultant value is up else show-down direction.

'=if (current-value>200, show-down, if (current value<-200, show-up, if (current-value>previous-value, show-up, else show-down))

The whole calculation of all these technical indicators prepared input value for the prediction model in the form of (0, 1) or binary-value. Our main focus in this research work is to find-out tomorrow stock market movement is up else show down-movement of stock market.

6. Experiments and result

The PSE data-set are collected for the experiment, the average trading days is about 4656. A new approach for stock market prediction which used RIDOR as a classifier is proposed. The proposed RIDOR base approach compared to existing classification techniques such as JRIP, Random forest, SVM and Naïve Bayes

6.1. Performance measure

To evaluate and validate the proposed model different performance measure are selected such as F-measure, precision and accuracy.

Precision formula:

\[ \text{Precision} = \frac{tp}{tp+fp} \]  
(1)

F-measure formula:

\[ \text{F-measure} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \]  
(2)

6.2. Result

The RIDOR technique show best result in the term of accuracy. The data preparation gives more help in implementation. The proposed technique show best accuracy of 89.62 percent. The proposed techniques compared with rule base classifier as well as machine learning techniques. Fig. 3 shows the result of performance measure of precision and F-measure.
7. Conclusion and future work

The stock-market prediction is important task for investors which have to be done before investing in the stock market. The accurate prediction of stock market gives profitable environment to investor. The stock market-prediction is complicated task because uncertainty involves which change the market (Ups/down). A new approach for stock market prediction which used RIDOR as a classifier is proposed. The proposed RIDOR base approach achieved 90% accuracy for stock market prediction while other classifiers such as JRIP, random forest, SVM, Naïve-Bayes achieved 89%, 88%, 89%, 86% respectively.

![Fig. 4: Accuracy of proposed technique](image)

### Table 1: Result of all comparison classifier

<table>
<thead>
<tr>
<th></th>
<th>RIDOR</th>
<th>JRIP</th>
<th>Random Forest</th>
<th>SVM</th>
<th>Naïve-Bayes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>0.914</td>
<td>0.878</td>
<td>0.889</td>
<td>0.895</td>
<td>0.86</td>
</tr>
<tr>
<td>Recall</td>
<td>0.89</td>
<td>0.874</td>
<td>0.889</td>
<td>0.894</td>
<td>0.86</td>
</tr>
<tr>
<td>Accuracy</td>
<td>90</td>
<td>87</td>
<td>89</td>
<td>89</td>
<td>86</td>
</tr>
</tbody>
</table>

References


