CHAPTER 6

FINDINGS, CONCLUSIONS AND SCOPE
FOR FUTURE RESEARCH

6.1 INTRODUCTION

This research work is an empirical investigation on predictability of both global and Indian stock markets in trying to predict select global indices using latest data mining tools such as k-NN, ANN, SVM, and decision tree models. Two different aspects of predictability of the markets – both forecasting and classification – are explored by employing forecasting and classifier versions of the above mentioned algorithms. Predictability of developed markets is compared with emerging markets. This study attempts to evaluate the efficacy of various forecasting and classifier models as well. Predictability of Indian markets is explored in depth to find out how predictability varies with different time frames. The study also tries to evaluate the impact of global cues on Indian stock markets and its predictability. It applies a new predictor set comprising of major global cues for predictive models and their performance is compared with that of models with technical indicator predictor set. By applying apriori algorithm, this study endeavors to identify frequently occurring patterns involving global cues and subsequently to generate association rules that would help predict the movement of Indian markets.
6.2 PREDICTABILITY OF GLOBAL MARKETS

6.2.1 Forecasting Models

Predictability among the global markets using forecasting models is compared using mean absolute percentage error and root mean squared percentage error.

- The highest forecasting error of MAPE is observed with RTSI (1.45%) and the lowest is found with GSPTSE (0.85%). The highest error of RMSPE occurs with SSE (2.25%) and the lowest is found with GSPTSE (1.15%). It is found that GSPTSE of Canada is the highly predictable market index and RTSI of Russia and SSEC of Shanghai are among the least predictable market indices.

- The average MAPE is found to be 0.97% for the indices of developed markets and it is found to be 1.28% for the emerging markets. The average RMSPE is found to be 1.34% for the indices of developed markets and it is found to be 1.85% for the emerging markets. Both the error metrics confirm that the indices of developed markets are more predictable than that of emerging nations.

- RMSPE of all the indices of developed markets are found to be below 1.5%, whereas all emerging nations are found to be above 1.5%. MAPE of all the indices of developed markets are found to be below 1.10%, whereas all emerging nations are found to be above 1.10%. NIKKEI is the only exception to this observation. This implies the fact that emerging markets are more turbulent and more difficult to forecast.
SSE and RTSI are among the lowest in terms of predictability and GSPFTE and FTSE are among highest in terms of predictability.

Indian indices of NSE NIFTY and BSE SENSEX exhibit lowest forecasting error among the emerging nations while CAC40 and DAX are the lowest among developed nations in terms of predictability.

6.2.2 Classifier Models

Predictability of the global markets, for classifier models, is compared using hit ratio, error rate, and kappa statistics.

The highest hit ratio of 56% is observed with NYSE and NASDAQ and the lowest hit ratio of 52% is observed with IBOVESPA.

The average hit ratio of developed markets is found to be 55% and it is only 53% in the case of emerging nations. It implies that the predictability of developed markets is higher than that of the emerging markets.

All emerging markets, with the exception of BSE SENSEX, is found to be having hit ratios below 53.6%, whereas all developed nations are found to attain hit ratios more than 53.60%.

The lowest hit ratio among the developed markets is observed with FTSE (53%). The highest hit ratio among the emerging markets occurs with BSE SENSEX (55%).
With the hit ratios of 55% and 52%, BSE SENSEX and NSE NIFTY are among the most predictable and least predictable indices, respectively.

From the results of classifier models and forecasting models, it is observed that the predictability of developed markets is higher than that of emerging markets.

6.3 THE EFFICACY OF PREDICTIVE MODELS

6.3.1 Forecasting Models

The performance of the forecasting models are evaluated by correlation coefficient, Mean absolute error, root mean squared error, relative absolute error, and root relative squared error.

Out of the 13 trials conducted, correlation coefficient obtained with SVM’s forecast remains in the top position. The k-NN correlation remains the least in all the trials. By considering average correlation coefficient value, SVM models is the most reliable of all the predictive models with a score of 0.9963, closely followed by ANN predictive model with a score of 0.9962. Decision tree retains the third spot with a score of 0.9959 and k-NN remains as the least accurate model with a score of 0.9932.

Average of mean absolute errors obtained from the 13 trials indicates that SVM is the most accurate forecasting model as its average error is the minimum with a value of 140.994. ANN remains the second best model with the second least error value of 149.049. Decision tree stands third with a value of 161.9031 and k-NN remains as the least performing model with a huge error value of 192.1086.
Average of root mean squared errors obtained in the 13 trials shows that SVM is the most accurate forecasting model as its average is the minimum with a value of 188.078. ANN remains as the second best model with the second least error value of 198.0863. Decision tree stands third with a value of 209.3455 and k-NN remains as the least performing model with a huge error value of 246.5919.

Average of relative absolute errors obtained from 13 trials also indicates that SVM is the best forecasting model as its average is the minimum with a value of 7.63%. ANN remains as the second best model with the second least error value of 7.78%. Decision tree stands third with a value of 8.82% and k-NN remains as the least performing model with an error value of 10.9%.

Average of root relative squared errors obtained from the 13 trials, SVM is the accurate forecasting model as its average is the minimum with a value of 8.47%. ANN remains as the second best model with the second least error value of 8.63%. Decision tree stands third with a value of 9.34% and k-NN remains as the least performing model with an error value of 11.65%.

SVM emerges as the most accurate forecasting model as it outperforms all the other models in terms of all the performance metrics considered for the study. SVM is closely followed by ANN model in terms of all the performance metrics. Decision tree-based model stands third and k-NN is found to be the least accurate forecasting model.
6.3.2 Classifier Models

Performance of classifiers is evaluated by comparing hit ratios, error rates, and kappa statistics.

- Out of the 13 trials, SVM’s classifier achieves the top hit ratio in eight occasions and the average hit ratio of this model is 54%. Though, random forest classifier could achieve the top hit ratio in only one out of the 13 trials, consistent performance takes it to the second spot as its average hit ratio is 51.36%. Though ANN model could achieve top hit ratio in four occasions out of thirteen trials, its average hit ratio (51.15%) is marginally below that of random forest. The k-NN is found to be the least accurate model with the average score of 50%.

- SVM classifier could achieve top kappa statistics value in eight out of thirteen trials. The average kappa statistics value is the highest with 0.052. The average values of kappa statistics for ANN and random forest classifier models are 0.017 and 0.004, respectively. The k-NN classifier scores negative value as its average value is the lowest (-0.016).

- SVM is also found to be the most accurate classifier, as it outperforms all the other models in terms of hit ratio, error rate, and kappa statistics. Though random forests could achieve higher hit ratio than ANN classifier, its score in kappa statistics is much lower than that of the ANNs. The k-NN is found to be the least accurate forecasting model in terms of both error metrics.
Results confirm the emergence of SVM as the most reliable model for forecasting and classification. ANN emerges as the next best model for classification and forecasting and it is closely followed by random forest. The k-NN seems to be the least efficient in forecasting and classification.

6.4 PREDICTABILITY OF INDIAN MARKETS

6.4.1 Forecasting Models

Predictability of Indian markets is explored with the SVM forecasting model, with two different predictor sets of technical indicators and global cues, on four test datasets of long term (364 days), medium term (170 days), short term (50 days), and very short term (25 days).

- While trying to capture the variation of forecasting errors for NSE NIFTY, it is seen that the root mean squared error falls significantly from 78.48 (long term) to the lowest value of 48.23 (very short term) through 57.99 and 51.22. There is a gradual reduction in forecasting error as we move from long term forecasting to short term forecasting. The same pattern is observed with mean absolute error also, as it falls from 56.79 to 37.75 through 44.9 and 39.17. It implies that predictability of NSE NIFTY improves significantly from long term to short term.

- In the case of BSE SENSEX, it is observed that the root mean squared error falls from 222.33 (long term) to the lowest value of 136.99 (very short term) through 194.32 and 151.55. Here also, it is observed that there is a significant fall in the root mean squared errors from long term to short term. The same pattern is observed with mean absolute error also, as it falls from 163.19 to 106.23 through 155 and 114.91. Predictability
of BSE SENSEX also seems to improve significantly from long term to short term.

- Steady and significant fall in forecasting errors from long term to very short term proves beyond doubt that Indian markets are better predictable in short term than in long term.

6.4.2 Classifier Models

- Performance of classifiers is compared in terms of the hit ratio and kappa statistics. It is observed that there is a remarkable improvement in hit ratio for NSE NIFTY from 54% (long term) to an impressive 70% (very short term) through 56%. Predictability improves significantly from long term to short term. The same pattern is observed with kappa statistics of NSE NIFTY, as this validity measure ascends from 0.058 to 0.32. It is evident that NSE NIFTY is better predictable in the short term than in the long term.

- In case of BSE SENSEX also, there is a remarkable improvement in hit ratio as it rises from 57% to 75% through 66%. The same pattern is observed with kappa statistics of BSE SENSEX, as it surges from 0.138 to 0.4898. It is evident that BSE SENSEX is more predictable in the short term than in the long term.

- Steady and significant fall in the forecasting errors observed with forecasting models and the impressive rise of hit ratio from 50 to 70% observed with classifiers, confirm beyond doubt, Indian indices are better predictable in the short term than in the long term.
6.5 GLOBAL CUES VERSUS TECHNICAL INDICATORS AS PREDICTOR SETS

6.5.1 Forecasting Models

- The average mean absolute errors of the forecasting models with global cues and technical indicator predictors in the four trials conducted for NSE NIFTY are found to be 45.86 and 46.93, respectively. As global cues model makes less forecasting error, it helps to predict better.

- The average root mean squared errors of the forecasting models with global cues and technical indicator predictors in the four trials conducted for NSE NIFTY are found to be 60.15 and 61.75, respectively. As global cues model makes less forecasting error, it helps to predict better.

- The average mean absolute errors of the forecasting models with global cues and technical indicator predictors in the four trials conducted for BSE SENSEX are found to be 134.83 and 138.92, respectively. Here also global cues model makes less forecasting error.

- The average root mean squared errors of the forecasting models with global cues and technical indicator predictors in the four trials conducted for BSE SENSEX are found to be 176.29 and 183.56, respectively. Global cues model makes less forecasting error.

- It is observed that global cues predictor set outperforms the technical indicator predictor set for 13 times out of the total of 16 trials. Moreover, average values of the forecasting errors also confirm that the predictive models with global cues perform better by making less predictive errors.
6.5.2 Classifier Models

- The average hit ratios of the classifier with global cues and technical indicator predictors in the four trials conducted for NSE NIFTY are found to be 58.24% and 57.74%, respectively. As it is seen, global cues edge over technical indicators.

- The average kappa statistics of the classifier with global cues and technical indicator predictors in the four trials conducted for NSE NIFTY are found to be 0.1561 and 0.1521, respectively. Global cues, as is seen here, get a better validation score than technical indicators.

- The average hit ratios of the classifier with global cues and technical indicator predictors in the four trials conducted for BSE SENSEX are found to be 64% and 61%, respectively. As it is seen, global cues outperform technical indicators in terms of success rate.

- The average Kappa statistics of the classifier with global cues and technical indicator predictors in the four trials conducted for BSE SENSEX are found to be 0.2597 and 0.1829, respectively. Global cues, as is seen here, get a better validation score than technical indicators.
6.6 EFFECT OF GLOBAL CUES ON INDIAN INDICES

The major findings of the year-wise and overall correlation analysis between Indian indices and global cues are listed below:

- It is found that the year-wise variation of correlation coefficient between Indian index and global cues are identical with BSE SENSEX and NSE NIFTY. It implies that the influence of global cues on both the indices is almost alike both in magnitude and direction.

- The American stock markets are found to exhibit a high degree of positive correlation with the Indian stock markets consistently throughout the period of study. NASDAQ is more correlated with Indian markets than NYSE.

- FTSE of London Stock exchange also displays positive correlation with the Indian markets.

- NIKKEI’s association with Indian indices oscillate between positive and negative values. Overall, NIKKEI is found to have negative correlation with Indian indices.

- US dollar is found to have moderate negative correlation with Indian indices.

- The pound of United Kingdom is found to have fluctuating correlation with Indian indices. Overall, it has a slight negative correlation with Indian indices.

- Euro’s association also fluctuates between positive and negative values. Overall, it is found to have a weak positive correlation with Indian indices.
- Yen exhibits weak positive correlation with Indian indices.
- Gold price displays a fluctuating association over the years. As a whole, it has positive correlation with Indian indices.
- Brent oil exhibits a positive correlation with Indian markets.
- LIBOR displays oscillating correlation with Indian indices, but overall it is found to have a moderate negative correlation with Indian markets.

6.7 ASSOCIATION RULE MINING FOR INDIAN INDICES

Apriori algorithm generates 16 rules for NSE NIFTY at a minimum support and confidence levels of 0.1 and 0.8, respectively. It also generates 16 rules for NSE NIFTY at a minimum support and confidence levels of 0.15 and 0.75, respectively.

Similarly, Apriori algorithm generates nine rules for BSE SENSEX at a minimum support and confidence levels of 0.1 and 0.8, respectively. It also generates 12 rules for BSE SENSEX at a minimum support and confidence levels of 0.15 and 0.75, respectively.

Support is a measure of frequency of association, confidence is a measure of strength of the association and lift is the measure of correlation between antecedent and consequent.

Though there are 32 distinct rules generated for NSE NIFTY and 21 rules for BSE SENSEX, many of the antecedents are common. The common association rules for both BSE SENSEX and NSE NIFTY with a minimum support of 0.1 and minimum confidence of 0.8, sorted in descending order of their lift value is given in Tables 6.1 and 6.2.
### Table 6.1 Common association rules for Indian indices (NSE NIFTY)

<table>
<thead>
<tr>
<th>Rule #</th>
<th>Antecedent</th>
<th>Consequent</th>
<th>Lift</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NIKKEI(B)=1, FTSE(B)=1, US$(B)=0, Yen(B)=0, RBRTE(B)=1</td>
<td>NSE(B)=1 162</td>
<td>8.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conf:(0.83)</td>
</tr>
<tr>
<td>2</td>
<td>NASDAQ(B)=1, FTSE(B)=1, US$(B)=0, Pound(B)=0</td>
<td>NSE(B)=1 163</td>
<td>7.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conf:(0.82)</td>
</tr>
<tr>
<td>3</td>
<td>NIKKEI(B)=1, FTSE(B)=1, Euro(B)=0, Yen(B)=0</td>
<td>NSE(B)=1 164</td>
<td>7.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conf:(0.8)</td>
</tr>
<tr>
<td>4</td>
<td>NIKKEI(B)=1, FTSE(B)=1, Pound(B)=0, Yen(B)=0</td>
<td>NSE(B)=1 173</td>
<td>7.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conf:(0.83)</td>
</tr>
<tr>
<td>5</td>
<td>NIKKEI(B)=1, FTSE(B)=1, US$(B)=0, Pound(B)=0</td>
<td>NSE(B)=1 171</td>
<td>7.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conf:(0.82)</td>
</tr>
<tr>
<td>6</td>
<td>NASDAQ(B)=1, NIKKEI(B)=1, FTSE(B)=1, US$(B)=0, Yen(B)=0</td>
<td>NSE(B)=1 173</td>
<td>7.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conf:(0.81)</td>
</tr>
<tr>
<td>7</td>
<td>NASDAQ(B)=1, FTSE(B)=1, US$(B)=0, RBRTE(B)=1</td>
<td>NSE(B)=1 180</td>
<td>7.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conf:(0.81)</td>
</tr>
<tr>
<td>8</td>
<td>NIKKEI(B)=1, FTSE(B)=1, US$(B)=0, RBRTE(B)=1</td>
<td>NSE(B)=1 195</td>
<td>6.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conf:(0.82)</td>
</tr>
<tr>
<td>9</td>
<td>NASDAQ(B)=1, NIKKEI(B)=1, FTSE(B)=1, US$(B)=0</td>
<td>NSE(B)=1 205</td>
<td>6.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conf:(0.81)</td>
</tr>
</tbody>
</table>

### Table 6.2 Common association rules for Indian indices (BSE SENSEX)

<table>
<thead>
<tr>
<th>Rule #</th>
<th>Antecedent</th>
<th>Consequent</th>
<th>Lift</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NIKKEI(B)=1, FTSE(B)=1, US$(B)=0, Yen(B)=0, RBRTE(B)=1</td>
<td>BSE(B)=1 161</td>
<td>8.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conf:(0.82)</td>
</tr>
<tr>
<td>2</td>
<td>NASDAQ(B)=1, FTSE(B)=1, US$(B)=0, Pound(B)=0</td>
<td>BSE(B)=1 161</td>
<td>8.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conf:(0.81)</td>
</tr>
<tr>
<td>3</td>
<td>NIKKEI(B)=1, FTSE(B)=1, Euro(B)=0, Yen(B)=0</td>
<td>BSE(B)=1 164</td>
<td>7.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conf:(0.8)</td>
</tr>
<tr>
<td>4</td>
<td>NIKKEI(B)=1, FTSE(B)=1, US$(B)=0, Pound(B)=0</td>
<td>BSE(B)=1 168</td>
<td>7.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conf:(0.81)</td>
</tr>
<tr>
<td>5</td>
<td>NIKKEI(B)=1, FTSE(B)=1, Pound(B)=0, Yen(B)=0</td>
<td>BSE(B)=1 168</td>
<td>7.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conf:(0.8)</td>
</tr>
<tr>
<td>6</td>
<td>NASDAQ(B)=1, NIKKEI(B)=1, FTSE(B)=1, US$(B)=0, Yen(B)=0</td>
<td>BSE(B)=1 172</td>
<td>7.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conf:(0.8)</td>
</tr>
<tr>
<td>7</td>
<td>NASDAQ(B)=1, FTSE(B)=1, US$(B)=0, RBRTE(B)=1</td>
<td>BSE(B)=1 182</td>
<td>7.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conf:(0.82)</td>
</tr>
<tr>
<td>8</td>
<td>NIKKEI(B)=1, FTSE(B)=1, US$(B)=0, RBRTE(B)=1</td>
<td>BSE(B)=1 194</td>
<td>6.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conf:(0.82)</td>
</tr>
<tr>
<td>9</td>
<td>NASDAQ(B)=1, NIKKEI(B)=1, FTSE(B)=1, US$(B)=0</td>
<td>BSE(B)=1 202</td>
<td>6.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>conf:(0.8)</td>
</tr>
</tbody>
</table>
Though there are 32 rules generated for NSE NIFTY, it is to be noted that all the rules, except one, are antecedents for rise in NSE NIFTY. Only one antecedent is generated for the fall in NSE NIFTY which is given below:

\[
\begin{align*}
\text{NASDAQ(B)} &= 0 \\
\text{NIKKEI(B)} &= 0 \\
\text{FTSE(B)} &= 0 \\
\text{RBRTE(B)} &= 0 \\
\text{NSE(B)} &= 0
\end{align*}
\]

\Rightarrow \quad \text{conf:} (0.8) \quad \text{lift:} (7.8)

This is the case observed with BSE SENSEX also. Though there are 21 rules generated, all but one are the antecedents for the rise in BSE SENSEX. Only one antecedent is generated for the fall in BSE SENSEX.

\[
\begin{align*}
\text{NIKKEI(B)} &= 0 \\
\text{FTSE(B)} &= 0 \\
\text{US$(B)$} &= 1 \\
\text{BSE(B)} &= 0
\end{align*}
\]

\Rightarrow \quad \text{conf:} (0.77) \quad \text{lift:} (5.10)

This indicates that not much patterns are generated that would help predict the bear trend of the indices and it implies that it is more difficult to predict the bear trend than the bull trend.

6.8 CONCLUSION

This empirical study is an attempt to measure the predictability of Indian and Global indices using the data mining approach. By employing four popular data mining tools predictability of the indices are measured in terms of value and direction. Predictability of emerging and developed markets is compared. The study reveals that emerging markets are less predictable than the developed markets. The results of the study prove the emergence of SVM as the most accurate forecasting and classifier model. Term-wise analysis of the predictability on Indian indices shows that markets are better predictable in short term than in the long term. The impressive growth in hit ratios of the classifiers and drastic reduction in forecasting errors, as one moves from long
term to short term prediction, is observed in the results. The study also highlights the effect of various global cues on Indian indices. These global cues, when used as a predictor set for predictive models, in lieu of the set of technical indicators, is found to result in a better predictive performance of the models. Association rule mining help identify patterns among the global cues that would result in bullish and bearish trends in the Indian indices. These association rules would be of immense help in predicting the trend of Indian markets. Thus this study throws more light on the dynamics and behavior of Indian indices with respect to developed and emerging stock markets.

6.9 SCOPE FOR FUTURE RESEARCH

The present study on the predictability of Indian and global stock indices using a data mining approach has provided many useful insights on the behavior of stock indices in the developed and emerging markets. The association between Indian stock indices and global cues opens more avenues to ponder as to how the other stock indices move with global cues. There is a wide scope for association rule mining with specific reference to individual stock indices. Market-wise association rules can be drawn to understand the association between the markets and global cues. Association rule mining may be extended for portfolio of stocks, mutual funds, and derivatives of market segments.

The present study is limited to select global indices and indices of BRIC nations alone. Further research on other global emerging markets and smaller markets can be carried out in depth using the above techniques to draw finer understanding on the predictability of the markets. Only select data mining tools have been chosen for the study as they have been proved to be robust and reliable in the past studies. However, other tools like Fuzzy tools and GA may be employed to get better results. Hybrid models may also
be employed to get new insights into the predictability of the stock index movements.

Technical analysis tools and global cues are the major input factors considered for the present study. Funke (2004) studied about the impact of the relationship between private consumption and stock markets for emerging markets with a cross-sectional time series data set. Estimates provided some evidence for a small, but statistically significant, link between private consumption growth and stock returns that are in line with the existence of a stock market wealth effect. Future studies can be aimed at studying the impact of monetary policy on the stock markets. How monetary policy can be designed in such a way that it has an influence on the stock market in emerging markets can be a very interesting study.

Only classification, prediction and association approaches of data mining are employed in the present study. Other aspects such as clustering and deeper pattern recognition may be applied to financial time series to draw more insights. Fuzzy data analysis also has received much attention for research in recent times. Finally, future research should consider the trading simulation under the scenarios of stock dividends, transaction costs, and individual-tax brackets to replicate the realistic investment practices.