Detection of Sentiment in CNX Nifty – An Investigative Attempt Using Probabilistic Neural Network

Bikramaditya Ghosh

Asst. Prof. ISME, Bangalore,

bikram77777@gmail.com

Dr. Padma Srinivasan

Assoc. Prof. Christ Univ., Bangalore,

bpadmasrinivasan@yahoo.co.in

Abstract

CNX Nifty has always been a remarkable benchmark for the Indian Capital Markets. The common wisdom believes that the prestigious Index is driven predominantly by FII cash inflow. Since another common belief states that FIIs check valuation and invest on the basis of valuation rather than on gut feel, so in a way CNX Nifty movement should ideally are predicted by P/E, P/B &Div. Yield etc. The entire world had witnessed a stellar rally in the Indian bourses between 3rd Jan 2007 to 2nd Jan 2008; that is the reason why that same time frame has been chosen in this study. Probabilistic neural Network (PNN) method is used this study, as it could replicate the human "Neuron" thinking quite well. This method allows the researcher with faster and accurate results. During that time zone FIIs Net Buys in Indian Securities were ₹ 9,480,196 MLN and the Net Sales were ₹ 8,389,304 MLN. So, effectively ₹ 1,090,892 MLN was Net investment. This was an all-time record that point of time.[1] In 2007 beginning FIIs were holding almost 16.3% of CNX Nifty. By the end of 2007 to the starting of 2008 they were holding 16.05% of the National Stock Exchanges' bellwether NSE 50 or Nifty. That clearly hints that despite reaching extraordinarily high valuation in Nifty, FIIs have not diluted their position. On the contrary they were holding or in other words consolidating their position in Nifty. At the same time, other participants too play a significant role.

---

[1] http://www.nseindia.com/content/us/ismr2011ch7.pdf

Objective & Problem Statement

The cardinal objective is to portray a period of an epic battle between two counterparties, sentiment vs. technical. In other words, Indian bourses started to witness steady to high capital inflow, with a hope of India shining along with a magnanimous boom in the Dalal Street. Investors and market participants irrespective of their investment knowledge and wisdom got driven by the age old greed & fear mechanism. Even negative information about the stocks were neglected and looked down upon. Optimism was riding a steep north bound direction.Though it was apparent that markets were following sentiments to a large extent during this phase of 2007-08, the major debate stands as" whether it is true that the investor groups got subdivided due to the presence of sentiment, despite their similar group characteristics!!" In a simple word, an HNI with a sound knowledge of investment vehicles suddenly wants to take part in that unprecedented bull rally, even at the cost of his asset allocation, which was originally decided by himself. FIIs with sound information and analysis base suddenly started to park money around Oct-November 2007.

Methodology

PNN, a Bayesian Neural Network with a high level of accuracy has been put to use in this study. It has been used as three input layers of three X variables (namely P/E, P/B & Div. Yield) having 250 cases of report (or observations during that 365 Days), and going through 10 neuron layers to come to form behaviorally similar groups. Time zone under the study starts from 3rd January 2007 and ends at 2nd January 2008. It has to be noted here that the crash in Indian securities started just after two days from that (i.e. 4th January 2008). The structure of a PNN directly reflects the Bayes criterion applied to the Parzen estimation method as it is a neural implementation of the Parzen windows. PNN however has certain disadvantages too, especially when it is used on a gigantic data base. It starts producing noise. However in a relatively small database, it works as an efficient and accurate tool.

Outcome

Neural Network Bayesian Classifier
Classification factor: Close
Input factors:

P/E
P/B
Div. Yield
Prior probabilities: non-informative
Error costs: equal for all classes

Number of cases in training set: 250
Number of cases in validation set: 0

Spacing parameter used: 0.0 (optimized by jackknifing during training)
Training Set
Table 1

Close	Members	Percent Correctly Classified
3576.5	1	100.0
3608.55	1	100.0
3626.85	1	100.0
3633.6	1	100.0
3641.1	1	100.0
3643.6	1	100.0
3655.65	1	100.0
3678.9	1	100.0
3690.65	1	100.0
3697.6	1	100.0
3718	1	100.0
3726.75	1	100.0
3733.25	1	100.0
3734.6	1	100.0
3745.3	1	100.0
3752	1	100.0
3761.1	1	100.0
3761.65	1	100.0
3764.55	1	100.0
3770.55	1	100.0
3798.1	1	100.0
3811.2	1	100.0
3819.95	1	100.0
3821.55	1	100.0
3829.85	1	100.0
3843.5	1	100.0
3848.15	1	100.0
3850.3	1	100.0
3861.05	1	100.0
3862.65	1	100.0
3875.9	1	100.0
3893.9	1	100.0
3911.4	1	100.0
3917.35	1	100.0
3933.4	1	100.0
3938.95	1	100.0
3942	1	100.0
3942.25	1	100.0
3983.4	1	100.0
3984.95	1	100.0
3988.8	1	100.0
3997.65	1	100.0
4011.6	1	100.0
4013.35	1	100.0
4024.05	1	100.0
4040	1	100.0
4044.55	1	100.0
4047.1	1	100.0
4052.45	1	100.0
4058.3	1	100.0
4066.1	1	100.0
4066.8	1	100.0
4074.9	1	100.0
4076.45	1	100.0
4076.65	1	100.0
4077	1	100.0
4078.4	1	100.0
4079.3	1	100.0
4080.5	1	100.0
4082.7	1	100.0
4083.5	1	100.0
4083.55	1	100.0
4085.1	1	100.0
4087.9	1	100.0
4089.9	1	100.0
4090.15	1	100.0
4096.2	1	100.0
4102.45	1	100.0
4106.95	1	100.0
4108.05	1	100.0
4109.05	1	100.0
4111.15	1	100.0
4113.05	1	100.0
4114.95	1	100.0
4117.35	1	100.0
4120.3	1	100.0
4124.45	1	100.0
4134.3	1	100.0

4137.2	1	100.0	4263.95	1	100.0	4496.75	1	100.0
4141.8	1	100.0	4267.05	1	100.0	4496.85	1	100.0
4145	1	100.0	4267.4	1	100.0	4497.05	1	100.0
4145.6	1	100.0	4278.1	1	100.0	4499.55	1	100.0
4146.2	1	100.0	4282	1	100.0	4504.55	1	100.0
4147.1	1	100.0	4284.65	1	100.0	4507.85	1	100.0
4147.7	1	100.0	4285.7	1	100.0	4509.5	1	100.0
4150.85	1	100.0	4293.25	1	100.0	4512.15	1	100.0
4153.15	1	100.0	4295.8	1	100.0	4518	1	100.0
4155.2	1	100.0	4297.05	1	100.0	4518.6	1	100.0
4164.55	1	100.0	4302.6	1	100.0	4528.85	1	100.0
4167.3	1	100.0	4313.75	1	100.0	4528.95	1	100.0
4170	1	100.0	4318.3	1	100.0	4546.2	1	100.0
4170.95	1	100.0	4320.7	1	100.0	4562.1	1	100.0
4171.45	1	100.0	4333.35	1	100.0	4566.05	1	100.0
4177.85	1	100.0	4339.5	1	100.0	4588.7	1	100.0
4178.6	1	100.0	4345.85	1	100.0	4619.35	1	100.0
4179.5	1	100.0	4353.95	1	100.0	4619.8	1	100.0
4183.5	1	100.0	4356.35	2	100.0	4620.75	1	100.0
4187.4	1	100.0	4357.55	1	100.0	4732.35	1	100.0
4190.15	1	100.0	4359.3	2	0.0	4747.55	1	100.0
4195.9	1	100.0	4370.2	1	100.0	4837.55	1	100.0
4198.25	1	100.0	4373.65	1	100.0	4932.2	1	100.0
4204.9	1	100.0	4384.85	1	100.0	4938.85	1	100.0
4209.05	1	100.0	4387.15	1	100.0	4940.5	1	100.0
4214.3	1	100.0	4401.55	1	100.0	5000.55	1	100.0
4214.5	1	100.0	4403.2	1	100.0	5021.35	1	100.0
4215.35	1	100.0	4406.05	1	100.0	5068.95	1	100.0
4219.55	1	100.0	4412.3	1	100.0	5085.1	1	100.0
4223.4	1	100.0	4419.4	1	100.0	5184	1	100.0
4224.25	1	100.0	4440.05	1	100.0	5185.85	1	100.0
4246.2	1	100.0	4445.2	1	100.0	5208.65	1	100.0
4248.15	1	100.0	4446.15	1	100.0	5210.8	1	100.0
4248.65	1	100.0	4462.1	1	100.0	5215.3	1	100.0
4249.65	1	100.0	4464	1	100.0	5327.25	1	100.0
4252.05	1	100.0	4474.75	1	100.0	5351	1	100.0
4256.55	1	100.0	4475.85	1	100.0	5428.25	1	100.0
4259.4	1	100.0	4479.25	1	100.0	5441.45	1	100.0
4260.9	1	100.0	4494.65	1	100.0	5473.7	1	100.0

5496.15	1	100.0		5742.3	1	100.0		5932.4	1	100.0
5519.35	1	100.0		5751.15	1	100.0		5937.9	1	100.0
5524.85	1	100.0		5762.75	1	100.0		5940	1	100.0
5559.3	1	100.0		5766.5	1	100.0		5954.7	1	100.0
5561.05	1	100.0		5777	1	100.0		5960.6	1	100.0
5568.95	1	100.0		5780.9	1	100.0		5974.3	1	100.0
5608.6	1	100.0		5782.35	1	100.0		5985.1	1	100.0
5617.1	1	100.0		5786.5	1	100.0		6047.7	1	100.0
5617.55	1	100.0		5847.3	1	100.0		6058.1	1	100.0
5634.6	1	100.0		5858.35	1	100.0		6070.75	1	100.0
5663.25	1	100.0		5865	1	100.0		6079.7	1	100.0
5668.05	1	100.0		5866.45	1	100.0		6081.5	1	100.0
5670.4	1	100.0		5868.75	1	100.0		6097.25	1	100.0
5695.4	1	100.0		5900.65	1	100.0		6138.6	1	100.0
5698.15	1	100.0		5905.9	1	100.0		6144.35	1	100.0
5698.75	1	100.0		5906.85	1	100.0		6159.3	1	100.0
5702.3	1	100.0		5907.65	1	100.0		6179.4	1	100.0
5731.7	1	100.0		5912.1	1	100.0		Total	**250**	**99.2**

This procedure uses a probabilistic neural network (PNN) to classify cases into different Close, based on 3 input variables of the 250 cases in the training set, 99.2% were correctly classified by the network.

Table 1.1

Row	Actual Group	Nearest Neighbor	Nearest Distance	2nd Nearest Neighbor	2nd Nearest Distance
124	4359.3	4357.55*	0.0000105952	4353.95	0.0000711042
163	4359.3	4320.7*	0.000332472	4302.6	0.000440044

Percent of training cases correctly classified: **99.20%**

Table 1.2

Close	Prior Probability	Error Cost								
3576.5	0.0040	1.0		3608.55	0.0040	1.0		3643.6	0.0040	1.0
				3626.85	0.0040	1.0		3655.65	0.0040	1.0
				3633.6	0.0040	1.0		3678.9	0.0040	1.0
				3641.1	0.0040	1.0		3690.65	0.0040	1.0

3697.6	0.0040	1.0	4052.45	0.0040	1.0	4155.2	0.0040	1.0
3718	0.0040	1.0	4058.3	0.0040	1.0	4164.55	0.0040	1.0
3726.75	0.0040	1.0	4066.1	0.0040	1.0	4167.3	0.0040	1.0
3733.25	0.0040	1.0	4066.8	0.0040	1.0	4170	0.0040	1.0
3734.6	0.0040	1.0	4074.9	0.0040	1.0	4170.95	0.0040	1.0
3745.3	0.0040	1.0	4076.45	0.0040	1.0	4171.45	0.0040	1.0
3752	0.0040	1.0	4076.65	0.0040	1.0	4177.85	0.0040	1.0
3761.1	0.0040	1.0	4077	0.0040	1.0	4178.6	0.0040	1.0
3761.65	0.0040	1.0	4078.4	0.0040	1.0	4179.5	0.0040	1.0
3764.55	0.0040	1.0	4079.3	0.0040	1.0	4183.5	0.0040	1.0
3770.55	0.0040	1.0	4080.5	0.0040	1.0	4187.4	0.0040	1.0
3798.1	0.0040	1.0	4082.7	0.0040	1.0	4190.15	0.0040	1.0
3811.2	0.0040	1.0	4083.5	0.0040	1.0	4195.9	0.0040	1.0
3819.95	0.0040	1.0	4083.55	0.0040	1.0	4198.25	0.0040	1.0
3821.55	0.0040	1.0	4085.1	0.0040	1.0	4204.9	0.0040	1.0
3829.85	0.0040	1.0	4087.9	0.0040	1.0	4209.05	0.0040	1.0
3843.5	0.0040	1.0	4089.9	0.0040	1.0	4214.3	0.0040	1.0
3848.15	0.0040	1.0	4090.15	0.0040	1.0	4214.5	0.0040	1.0
3850.3	0.0040	1.0	4096.2	0.0040	1.0	4215.35	0.0040	1.0
3861.05	0.0040	1.0	4102.45	0.0040	1.0	4219.55	0.0040	1.0
3862.65	0.0040	1.0	4106.95	0.0040	1.0	4223.4	0.0040	1.0
3875.9	0.0040	1.0	4108.05	0.0040	1.0	4224.25	0.0040	1.0
3893.9	0.0040	1.0	4109.05	0.0040	1.0	4246.2	0.0040	1.0
3911.4	0.0040	1.0	4111.15	0.0040	1.0	4248.15	0.0040	1.0
3917.35	0.0040	1.0	4113.05	0.0040	1.0	4248.65	0.0040	1.0
3933.4	0.0040	1.0	4114.95	0.0040	1.0	4249.65	0.0040	1.0
3938.95	0.0040	1.0	4117.35	0.0040	1.0	4252.05	0.0040	1.0
3942	0.0040	1.0	4120.3	0.0040	1.0	4256.55	0.0040	1.0
3942.25	0.0040	1.0	4124.45	0.0040	1.0	4259.4	0.0040	1.0
3983.4	0.0040	1.0	4134.3	0.0040	1.0	4260.9	0.0040	1.0
3984.95	0.0040	1.0	4137.2	0.0040	1.0	4263.95	0.0040	1.0
3988.8	0.0040	1.0	4141.8	0.0040	1.0	4267.05	0.0040	1.0
3997.65	0.0040	1.0	4145	0.0040	1.0	4267.4	0.0040	1.0
4011.6	0.0040	1.0	4145.6	0.0040	1.0	4278.1	0.0040	1.0
4013.35	0.0040	1.0	4146.2	0.0040	1.0	4282	0.0040	1.0
4024.05	0.0040	1.0	4147.1	0.0040	1.0	4284.65	0.0040	1.0
4040	0.0040	1.0	4147.7	0.0040	1.0	4285.7	0.0040	1.0
4044.55	0.0040	1.0	4150.85	0.0040	1.0	4293.25	0.0040	1.0
4047.1	0.0040	1.0	4153.15	0.0040	1.0	4295.8	0.0040	1.0

4297.05	0.0040	1.0	4518.6	0.0040	1.0	5634.6	0.0040	1.0
4302.6	0.0040	1.0	4528.85	0.0040	1.0	5663.25	0.0040	1.0
4313.75	0.0040	1.0	4528.95	0.0040	1.0	5668.05	0.0040	1.0
4318.3	0.0040	1.0	4546.2	0.0040	1.0	5670.4	0.0040	1.0
4320.7	0.0040	1.0	4562.1	0.0040	1.0	5695.4	0.0040	1.0
4333.35	0.0040	1.0	4566.05	0.0040	1.0	5698.15	0.0040	1.0
4339.5	0.0040	1.0	4588.7	0.0040	1.0	5698.75	0.0040	1.0
4345.85	0.0040	1.0	4619.35	0.0040	1.0	5702.3	0.0040	1.0
4353.95	0.0040	1.0	4619.8	0.0040	1.0	5731.7	0.0040	1.0
4356.35	0.0040	1.0	4620.75	0.0040	1.0	5742.3	0.0040	1.0
4357.55	0.0040	1.0	4732.35	0.0040	1.0	5751.15	0.0040	1.0
4359.3	0.0040	1.0	4747.55	0.0040	1.0	5762.75	0.0040	1.0
4370.2	0.0040	1.0	4837.55	0.0040	1.0	5766.5	0.0040	1.0
4373.65	0.0040	1.0	4932.2	0.0040	1.0	5777	0.0040	1.0
4384.85	0.0040	1.0	4938.85	0.0040	1.0	5780.9	0.0040	1.0
4387.15	0.0040	1.0	4940.5	0.0040	1.0	5782.35	0.0040	1.0
4401.55	0.0040	1.0	5000.55	0.0040	1.0	5786.5	0.0040	1.0
4403.2	0.0040	1.0	5021.35	0.0040	1.0	5847.3	0.0040	1.0
4406.05	0.0040	1.0	5068.95	0.0040	1.0	5858.35	0.0040	1.0
4412.3	0.0040	1.0	5085.1	0.0040	1.0	5865	0.0040	1.0
4419.4	0.0040	1.0	5184	0.0040	1.0	5866.45	0.0040	1.0
4440.05	0.0040	1.0	5185.85	0.0040	1.0	5868.75	0.0040	1.0
4445.2	0.0040	1.0	5208.65	0.0040	1.0	5900.65	0.0040	1.0
4446.15	0.0040	1.0	5210.8	0.0040	1.0	5905.9	0.0040	1.0
4462.1	0.0040	1.0	5215.3	0.0040	1.0	5906.85	0.0040	1.0
4464	0.0040	1.0	5327.25	0.0040	1.0	5907.65	0.0040	1.0
4474.75	0.0040	1.0	5351	0.0040	1.0	5912.1	0.0040	1.0
4475.85	0.0040	1.0	5428.25	0.0040	1.0	5932.4	0.0040	1.0
4479.25	0.0040	1.0	5441.45	0.0040	1.0	5937.9	0.0040	1.0
4494.65	0.0040	1.0	5473.7	0.0040	1.0	5940	0.0040	1.0
4496.75	0.0040	1.0	5496.15	0.0040	1.0	5954.7	0.0040	1.0
4496.85	0.0040	1.0	5519.35	0.0040	1.0	5960.6	0.0040	1.0
4497.05	0.0040	1.0	5524.85	0.0040	1.0	5974.3	0.0040	1.0
4499.55	0.0040	1.0	5559.3	0.0040	1.0	5985.1	0.0040	1.0
4504.55	0.0040	1.0	5561.05	0.0040	1.0	6047.7	0.0040	1.0
4507.85	0.0040	1.0	5568.95	0.0040	1.0	6058.1	0.0040	1.0
4509.5	0.0040	1.0	5608.6	0.0040	1.0	6070.75	0.0040	1.0
4512.15	0.0040	1.0	5617.1	0.0040	1.0	6079.7	0.0040	1.0
4518	0.0040	1.0	5617.55	0.0040	1.0	6081.5	0.0040	1.0

6097.25	0.0040	1.0
6138.6	0.0040	1.0
6144.35	0.0040	1.0

6159.3	0.0040	1.0
6179.4	0.0040	1.0

Figure 1

Neural Network

| Input layer	Pattern layer	Summation layer	Output layer
(3 variables)	(250 cases)	(10 neurons)	(10 groups)
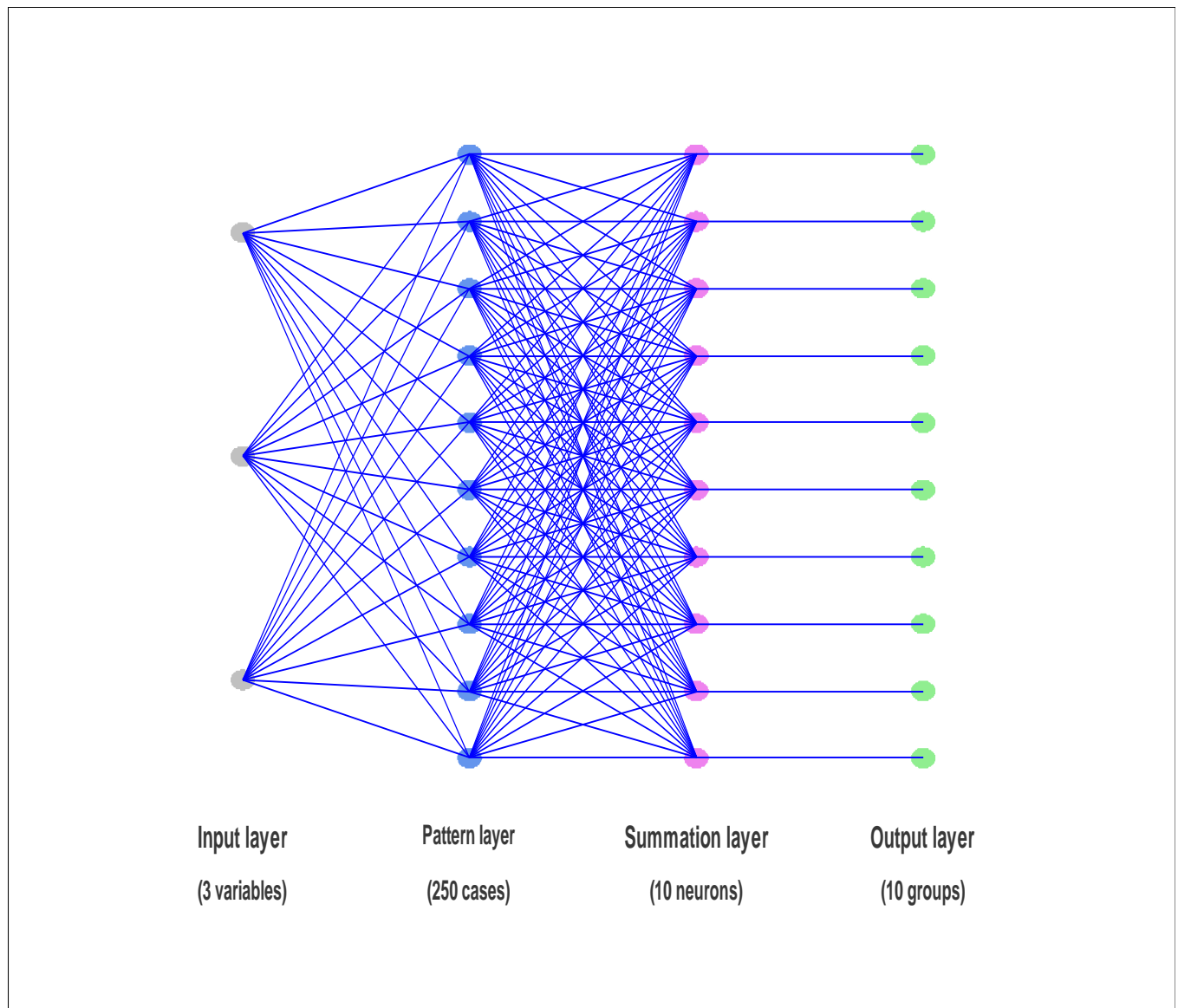
Stock Contour as a 3D Model

Figure 1.1
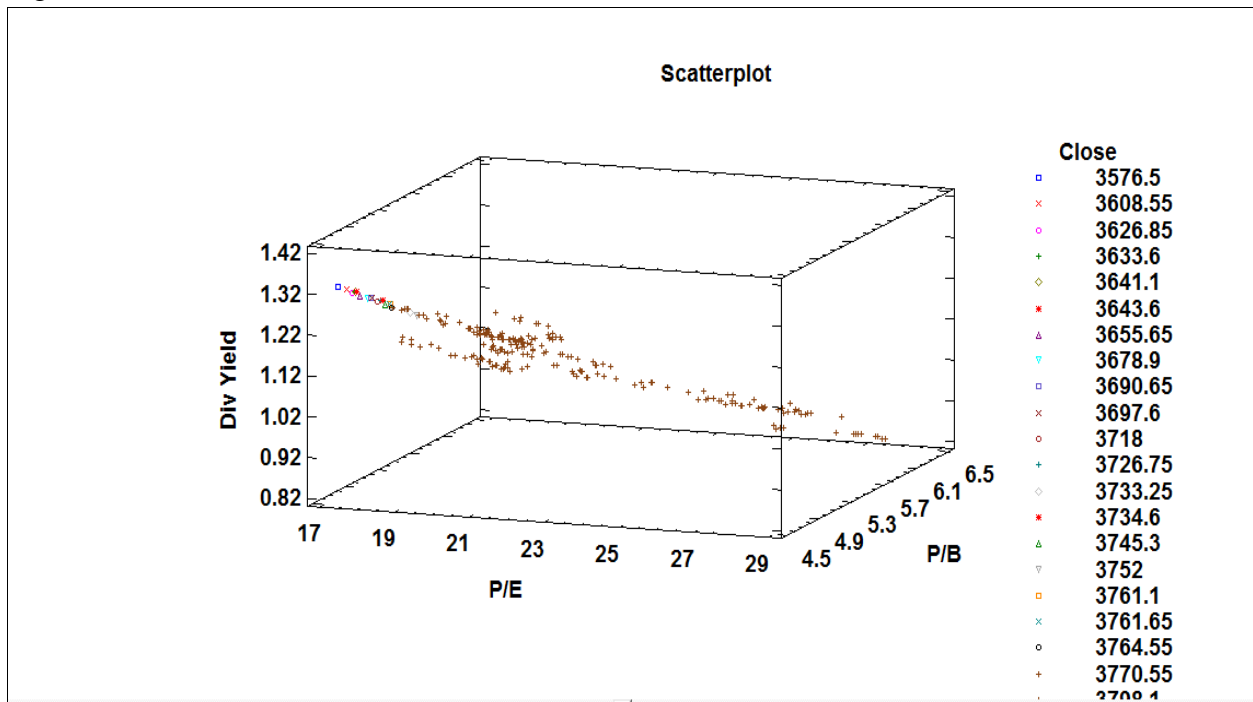
Conclusion & Interpretation

[1] The outcome clearly depicts that 1 crucial year of Indian Capital Market, the market participants were broadly divided into 10 groups of distinctly different buying of selling behaviour of CNX Nifty.

[2] Generally we have 5 investor groups or market participants in Indian Capital Markets, namely FII, DII, SME, HNI & Retail. However in this case we found that even sub classes became prominent. Could be completely opposite views within each of the market participant classes have further made it into two halves.

[3] In a simple way, may be at P/E level of 27.8, on 2nd January 2008, even some of the or could be majority of the HNIs started to have a view of coming out of the Index, as the valuations have reached euphoric levels with such an high expectations from the forward earnings side.

[4] As the prediction and classification accuracy is 99.2%, so it could easily be identified that the outcome is fairly correct. The judgment model is more is quite strong. Now had there

been no trace of sentiment or confidence factor, then the outcome groups should have been far lesser.

[5] In fact the groups should ideally be five. This clearly indicates that these groups too change group dynamics based on valuation data.

[6] One part of the group follow technical and the other follows hope, greed & fear mechanism. Substantiating the fact that Indian stock market was its euphoric highs during the said period, in fact the uninterrupted Bull rally that started way back in mid-2003, was at its pick and with valuations being stretched every now and then, was due for an overdue correction or could possibly a crash.

Limitations of the Study

1. This study is for a brief time zone of 365 Days with 250 trading days
2. This study is done only on CNX Nifty
3. This study is solely for the purpose of detection of investor sentiment
4. PNN is used, and not any other Gaussian Mixture Model

Scope for further research

1. Could have been done on S&P BSE Sensex or any other larger index
2. Time zone could be different and longer
3. Any other behaviour could be detected too
4. Any other model could be used

Key Take Away

The major outcome from this study depicts a craggy reality, which is possibly could be denoted as "Sentiment".Investor groups based on certain demographics & psychographics even behave in a clearly heterogeneous manner. This however is splitting the group even further in to various subgroups.

References

[1] Burrascano, P. (1991). Learning vector quantization for the probabilistic neural network. *IEEE Transactions on Neural Networks*, pp. 458-461.
[2] Fabio Ancona, A. M. (1998). Implementing probabilistic neural networks. *Neural Comput & Applic*, pp. 37-51.
[3] Ghosh, B. (2014). BSE 100 Market Capitalization follows Sentiment of Investors or Technical Methods-An analytical study. *SRJIS*, pp. 400-404.

[4]   Specht, D. F. (1990). "Probabilistic neural networks". . *Neural Networks 3*, pp. 109–118.

[5]   Stern, H. (1996). Neural Networks in Applied Statistics. *Technometrics*, pp. 205-214.

[6]   Tkacz, G. (2001). Neural Network Forecasting of Canadian GDP Growth. *International Journal of Forecasting*, pp. 57-69.