Stock Market Analysis With Machine Learning

An Investigative Report into It's Feasibility

What is the Stock Market?

- It is a market where people buy stakes of companies.
- It is regulated by SEBI, here in India and it backs the entire economy of India.

However, it is extremely volatile. Prices fluctuate between extremes. It is a near perfect representation of a random event generator.

Is there some pattern to the chaos?

What Does Stock Market Depend On?

- Supply
- Demand

These are the two factors solely controlling the stock market.

Supply and Demand are difficult to predict.
So, other indicators can be looked at.

In this case, the Indian Stock market can be investigated.

Indicators of Stock Market Prices

- Change in Consumer Price Index (CPI).
- Prices of non-equity assets like government bonds and gold.
- Prices of crude oil, iron, aluminum and wheat.
- Demographics and population distribution.
- Rainfall.
- Natural Disasters.

However, the most important data the stock market can be predicted with, which many of us use, is past and present index data.

Indexes and their historic datasets can give immense insights into the performance of a particular company, sector, country or economic region.

Indexes that may be considered

NIFTY and S&P BSE SENSEX

These are free-float market-weighted stock market indices of 30 well-established and financially sound companies listed on Bombay Stock Exchange. The 30 component companies which are some of the largest and most actively traded stocks, are representative of various industrial sectors of the Indian economy.

Indexes that may be considered

S&P BSE India Government Bond Index

It is designed to track the performance of local-currency denominated government and corporate bonds from India.

Indexes that may be considered

S&P BSE 100 and S&P BSE SENSEX 50

These indices track the performance of the top 100 and top 50 companies respectively, in India, across various sectors. These represent the performance of the biggest companies in the country.

Indexes that may be considered

S&P BSE Bharat 22 Index

It is designed to measure the performance of 22 select companies disinvested by the central government of India. It was formed relatively recently and gives a good insight into the non-nationalised large-cap and mid-cap companies.

A Machine Learning algorithm may figure out the order in all the chaos by looking at the datasets of external indicators and index performances. Daily, Monthly and Quarterly data between September, 2009 and August 2018 may be considered.

What happens if there is no order?

- It is entirely possible that no order may be found.
- In that case, information from the stock markets can be used in random number generators.

Finding a pattern will give tremendous insights into the act of investing into companies and will help people make better investments in better companies. This will add to the financial security of our country.

What We Will be Looking For?

We will be trying to answer the following problems with our software:

- Is the stock high risk or low risk?
- Is the sector high risk or low risk?
- Will the market be bullish or bearish in the next few months/years?
- Investment in which sector is profitable in the near and distant future?

What Datasets Will We Look At?

We will be looking at the Indian Stock Market. So, the following data will be considered:

- **Index data (Daily)** : NIFTY and S&P BSE SENSEX, S&P BSE Bharat 22 Index, S&P BSE 100 and S&P BSE SENSEX 50 and S&P BSE India Government Bond Index (from S&P BSE and NIFTY website).
- **Environmental Data (Monthly)** : Rainfall in all states in India (in centimeters).
- **Commodity Prices (Daily)** : Crude Oil (Not Seasonally Adjusted; USD per barrel; Source: Federal Bank of St. Louis, USA; global prices), Gold (Rupees per kilogram), Global Metal Index (Index 2016 = 100, Not Seasonally Adjusted; Source: Federal Bank of St. Louis, USA).
- **Exchange Rates (For G20 Nations or economic regions) (Monthly Data)** : National Currency per U.S. Dollar (Period average), Real Effective Exchange Rate Index. (Source: Principal Global Indicators (PGI) Database)
- **Price, Production and Labour (For G20 Nations or economic regions) (Monthly Data)** : Consumer Price Index (CPI) (% change from previous period), Industrial Production Index (% change from previous period). (Source: Principal Global Indicators (PGI) Database)

How Will the Data Look?

ASIA INDEX Pvt. Ltd.
A BSE and S&P DJI Venture

As of:	Aug 02, 2019				
	Data has been based at 100				
Effective date	Bond Index	S&P BSE SENSEX (TR)	S&P BSE 100 (TR)	S&P BSE Bharat 22 Index TR	S&P BSE SENSEX 50 TR
31/07/2009	100	100	100	100	100
03/08/2009	99.96	101.63	101.65	100.62	101.55
04/08/2009	99.87	101.04	101.08	100.08	101.01
05/08/2009	99.82	101.51	101.42	100.05	101.44
06/08/2009	99.42	99.02	99.04	97.85	99.09
07/08/2009	99.39	96.76	96.73	96.5	96.87
10/08/2009	99.35	95.8	95.57	95.49	95.96
11/08/2009	99.43	96.22	96.27	96	96.59
12/08/2009	99.56	95.87	96.08	95.27	96.33
13/08/2009	99.51	99.05	99.31	98.83	99.41
14/08/2009	99.49	98.37	98.73	99.15	98.79
17/08/2009	99.31	94.37	94.67	94.92	94.73
18/08/2009	99.11	96.02	96.26	97.1	96.27
19/08/2009	99.13	94.58	94.71	96.11	94.78
20/08/2009	99.05	95.87	95.98	97.48	96.08
21/08/2009	99.09	97.34	97.5	98.63	97.59
24/08/2009	98.67	99.82	99.86	100.65	99.97
25/08/2009	98.7	100.2	100.21	100.5	100.27

How Will the Data Look?

Consumer Price Index, All items, Percentage change over corresponding period previous year

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Australia	4.35	1.77	2.92	3.30	1.76	2.45	2.49	1.51	1.28	1.95
Austria	3.22	0.51	1.81	3.29	2.49	2.00	1.61	0.90	0.89	2.08
Belgium	4.49	-0.05	2.19	3.53	2.84	1.11	0.34	0.56	1.97	2.13
Brazil	5.68	4.89	5.04	6.64	5.40	6.20	6.33	9.03	8.74	3.45
Canada	2.37	0.30	1.78	2.91	1.52	0.94	1.91	1.13	1.43	1.60
China, P.R.: Hong Kong	4.30	0.58	2.31	5.28	4.06	4.32	4.44	3.00	2.41	1.48
China, P.R.: Mainland	5.93	-0.73	3.18	5.55	2.62	2.62	1.92	1.44	2.00	1.59
Denmark	3.42	1.30	2.31	2.76	2.40	0.79	0.56	0.45	0.25	1.15
Finland	4.07	0.00	1.18	3.42	2.81	1.48	1.04	-0.21	0.36	0.75
France	2.81	0.09	1.53	2.11	1.95	0.86	0.51	0.04	0.18	1.03
Germany	2.63	0.31	1.10	2.08	2.01	1.50	0.91	0.51	0.49	1.51
India	8.35	10.88	11.99	8.86	9.31	10.91	6.35	5.87	4.94	2.49
Indonesia	10.23	4.39	5.13	5.36	4.28	6.41	6.39	6.36	3.53	3.81
Ireland	4.06	-4.48	-0.92	2.56	1.70	0.51	0.18	-0.29	0.01	0.34
Italy	3.35	0.77	1.53	2.78	3.04	1.22	0.24	0.04	-0.09	1.23
Japan	1.38	-1.35	-0.72	-0.27	-0.05	0.35	2.76	0.79	-0.12	0.47
Korea, Republic of	4.67	2.76	2.94	4.03	2.19	1.30	1.27	0.71	0.97	1.94

After the preliminary stages of development, population distribution and general mood of the people and a separate randomness variable will be introduced to make predictions more informed.

After all this, we are targeting to get an accuracy of about 80%

The exact nature of the prediction algorithm is subject to change as more research is done on this topic.

Thank You!