

# Data Visualization and Prediction using Candlesticks Chart for Open Government DataSet

M.Uma, M.S.Abirami, S. Metilda Florence

**Abstract:** Open government data is a very helpful initiative to transfer the usefulness of data by putting all datasets in public domain. This data can be leveraged for back testing, decision making, forecast modelling, and risk assessment and so on. The Government dataset available in the portals (Data.gov.in) is only raw format, as the data is provided in raw format (.csv, .json etc.), it is hard to make much sense of it without pre-processing and filter. It requires some level of technical expertise to compile, filter, transform and visualize the data to make sense. In the proposed system implemented the mobile application to importing datasets from public domain sources and then plotting for user-understanding and friendly manner. As most of the goods prices move because of the supply-demand differences, the proposed Japanese candlestick charts used to predict/forecast patterns in short term and long term by capturing the change in trend using moving averages, pattern recognition methods. This forecasting model has been used in Japan to forecast wheat/grain prices with high level of accuracy. As most of the Small and Medium Business owners, farmers, Traders are not so tech-savvy it become a hurdle for them to get data and interpret metrics from it. By user friendly and interactive visual charts helps them to take quick decision making, risk assessment, certainty on outcomes. Hence, the proposed mobile application of candlesticks chart gives clear picture of the product price and also easily access anytime and anywhere.

**Keywords:** About four key words or phrases in alphabetical order, separated by commas.

## I. INTRODUCTION

Farmers, SME's, Businesses have to struggle hard to decide at what time and price to sell their stock to get highest return on capital employed. This has become much harder in today's marketplace where there are many stakeholders bidding for the commodity. As there many other factors affecting the price of the commodity, but at the end it's the price that seller gets that matters. To help this segment of people to take better decision by using data as a backing point of all decisions, government has started a digital initiative Open government data (data.gov.in) portal to provide all public data. As the data is provided in raw format (.csv, .json etc.), it is hard to make much sense of it without pre-processing and filter.

In this report, we have put forth a solution to this by processing the data provided by the government and using some widely adopted charting and pattern techniques to help

the farmers and SME's by providing them basic visualization through charts and also indicators which might help them make sense out of it. Here raw data is imported from data.gov.in portal, it is processed and filtered with

required datasets and then plotted. As the chart is plotted using candlesticks (which is more data rich than any other chart). Then we are using various algorithms to give buy/sell signals indicators to give signals on various predefined set of rules.

## II. LITERATURE SURVEY

Hongbing Ouyang, et.al.[1] to find out how high low of an underlying assets price movement carry more information than the traditional line, bar chart. Because of this data rich nature of candlesticks, candlestick charting is very helpful in analysing volatility, price trend, decision taking. Patterns and more developed techniques enhance candlestick charting for looking into the numbers and make more out of the given dataset. RenatoDe Donato, Delfina Malandrino, Giuseooina Palmieri, et.al.[2] the DataEt-Ecosystem Provider (DEEP), The aim of DEEP is to engage various stakeholders of the society to make more use of this available datasets and look into modelling better decision making techniques backed by this data. This can be achieved only by making it more accessible to masses, redefine use cases, better interface and adaptability. And by using DEEP architecture to process data on clouds to increase adaptability among low end systems. Chiung-Hon Leon Lee, Yi-Ching Liaw and Lindroos Hsu, et.al.[3] an approach to extract fuzzy candlestick patterns from a candlestick chart and feed the system with some well proven patterns with best risk reward ratio and to deploy it into stock markets terminals so that the system will screen the best trades for the investor and help him take back tested decisions or at its best automate the whole thing.

Yang Shulin and Hu Jieping, et.al.[4] Web services and mobile development increases the penetration of internet among masses and make them digitally connected and informed about the world and for their own welfare. Volley framework launched by google does a great job for mobile development but it leaves the web service users at the gates. Leo Rizky Julian and Friska Natalia, et.al. [5] This application was made keeping in mind the general personal modifications and demands from a custom system. It allows user to decide and customised a truly personal computer based on his/her requirement and can be tweaked with specification. This proves to be a better value for money and has better adaptability

**Revised Manuscript Received on July 05, 2019.**

M.Uma,  
M.S.Abirami,  
S. Metilda Florence

and use case than the traditional one size fits all. This allows users to configure and imagine their own systems to work on and get the best out of their systems, this also enables better engagement, efficiency and working of users.

### III. OPEN GOVERNMENT DATA (OGD) PLATFORM INDIA- DATASET

Data.gov.in - is an online platform an initiative of Government of India to provide the open data to any users shown in Fig 1 and dataset of wheat shown in table1. This website is mainly used by the different department of the government and by different organisation under the government. This portal provides the sets of datasets which have the information related to everything which comes under the government, it helps the common people or any organisation to check these data from the website for individual knowledge.

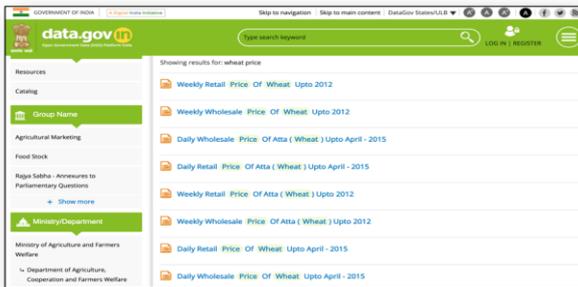


Figure1: Primary source of open government datasets (data.gov.in) web portal

Table 1: Wheat Dataset

Date	Centre_Name	Commodit Price	
25-11-1997	DELHI	Wheat	6.5
25-11-1997	SHIMLA	Wheat	6.2
25-11-1997	LUCKNOW	Wheat	4.9
25-11-1997	AHMEDABAD	Wheat	7.5
25-11-1997	BHOPAL	Wheat	6
25-11-1997	MUMBAI	Wheat	9
25-11-1997	JAIPUR	Wheat	6.5
25-11-1997	PATNA	Wheat	5.4
25-11-1997	BHUBANESHWAR	Wheat	7.5
25-11-1997	KOLKATA	Wheat	7
25-11-1997	GUWAHATI	Wheat	4.7
25-11-1997	SHILONG	Wheat	8
25-11-1997	AIZWAL	Wheat	5.5
25-11-1997	AGARTALA	Wheat	8.2
25-11-1997	BENGALURU	Wheat	9.5
25-11-1997	T.PURAM	Wheat	12
25-11-1997	CHENNAI	Wheat	11
25-11-1997	HYDERABAD	Wheat	8
26-11-1997	DELHI	Wheat	6.5
26-11-1997	SHIMLA	Wheat	7.6
26-11-1997	LUCKNOW	Wheat	4.9
26-11-1997	AHMEDABAD	Wheat	7.5
26-11-1997	BHOPAL	Wheat	5.5
26-11-1997	MUMBAI	Wheat	8.5
26-11-1997	JAIPUR	Wheat	6.5
26-11-1997	PATNA	Wheat	6.6
26-11-1997	BHUBANESHWAR	Wheat	7.5

#### A. System Architecture

System Architecture is a practical model that describes the structure and behaviour of complex systems. It comprises of all the components and gives the overview of the overall system shown in Fig 2

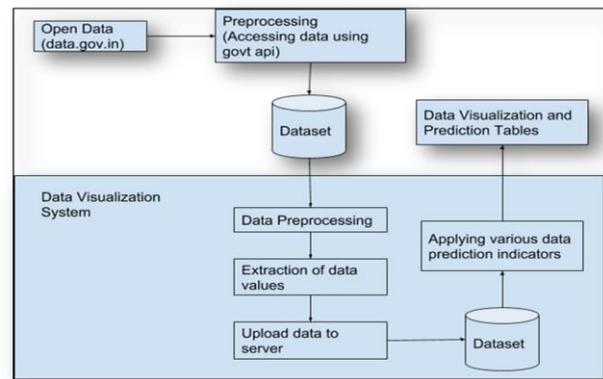


Figure 2: Architecture of Data Visualization and Prediction Model

#### Modules for Implementation

There are basically four main modules present in the application which are used for the open data visualization **A. Data Extraction from Data.gov.in**

All the required data is imported in raw format (.csv, .json etc.) from data.gov.in portal and then filtered according to the requirements and feed it to the system where visualisation/ charting is done according to demography and time frames. The data is downloaded by the admin itself from the open data sites and the same is stored on the database server and downloaded from the server to the user android operated mobile device. The user will be given the set of different datasets and it will be fetched using the internet connectivity of the device.

#### B. Data Cleaning

The data received from the data.gov.in contains of various features which need to be collected and the same is needed for the calculation of the candle sticks plotting parameters like open, close, high, low and these values are needed for applying the different data prediction indicators, which works on the large data set to produce results.

#### C. Features Extraction

Imported data from the portal cannot be used for the system, it requires some data manipulation according to the system's requirement. The CSV format of dataset not understandable by the end user. In order to avail the data to user in better understandable state, a candle stick chart was proposed which is widely used in financial market currently. It helps the end user to get four values at the single point of time. The four parameters are low, high, open and close. These parameters are needed to plot the candle stick chart which were passed to the chart plotting library.

#### D. Data Visualisation using Candlesticks

After extracting all required features from the dataset, it is required to plot using candlestick chart to visualise the datasets in more understandable and informative manner. To provide clear picture candle stick chart and line chart were used. To implement these charts, android library MPAndroidChart



used for plotting these chart for clear data visualization.

### E. Data Prediction

The data plotting graph produces the four different values at the particular point of the time. The same parameter can be used and applied in the different data prediction algorithm. So user can easily predict the price value of the product.

The visualization of Candle Stick chart showing the weekly price of wheat for Ranchi in Fig3. And Relative Strength Indicator RSI shown in Fig.4.



Figure 3: Candle stick chart -Weekly price of wheat in Ranchi.

Sr	Date	Clos	Chang	Gai	Los	AvgGai	AvgLos	R	Rr
122	29-05-98	7	0.00	0.00	0	0.01	-0.01	1.50	60.00
123	01-06-98	7	0.00	0.00	0	0.01	-0.01	1.50	60.00
124	02-06-98	7	0.00	0.00	0	0.01	-0.01	1.50	60.00
125	03-06-98	7	0.00	0.00	0	0.01	-0.01	1.50	60.00
126	04-06-98	7	0.00	0.00	0	0.01	-0.01	1.50	60.00
127	05-06-98	7	0.00	0.00	0	0.01	-0.01	1.50	60.00
128	08-06-98	7	0.00	0.00	0	0.01	-0.01	1.50	60.00
129	09-06-98	7	0.00	0.00	0	0.01	-0.01	1.50	60.00
130	10-06-98	6.5	-0.50	0	-0.50	0.01	-0.01	1.00	50.00
131	11-06-98	6.5	0.00	0.00	0	0.01	-0.01	1.00	50.00
132	12-06-98	6.5	0.00	0.00	0	0.01	-0.01	1.00	50.00
133	15-06-98	6.5	0.00	0.00	0	0.01	-0.01	1.00	50.00
134	16-06-98	6.5	0.00	0.00	0	0.01	-0.01	1.00	50.00
135	17-06-98	6.5	0.00	0.00	0	0.01	-0.01	1.00	50.00
136	18-06-98	6.5	0.00	0.00	0	0.01	-0.01	1.00	50.00
137	19-06-98	6.5	0.00	0.00	0	0.01	-0.01	1.00	50.00
138	22-06-98	6.5	0.00	0.00	0	0.01	-0.01	1.00	50.00
139	23-06-98	6.5	0.00	0.00	0	0.01	-0.01	1.00	50.00
140	24-06-98	6.5	0.00	0.00	0	0.01	-0.01	1.00	50.00
141	25-06-98	6.5	0.00	0.00	0	0.01	-0.01	1.00	50.00
142	26-06-98	6.5	0.00	0.00	0	0.01	-0.01	1.00	50.00
143	29-06-98	6.5	0.00	0.00	0	0.01	-0.01	1.00	50.00
144	30-06-98	6.5	0.00	0.00	0	0.01	-0.01	1.00	50.00

Figure 4: Relative Strength Indicator calculation in Ranchi (between 55-30 is considered average RSI)

### F. Implementation of Improved Hyperlocal Indicators for Prediction

Once the plotting is done, most recognised and followed charting algorithms/indicators used for plotting for various demography and time frame to get overview of the underlying and its performance. Here the different data indicators are used for decision making. The pivot point can be used to calculate support and resistance, second indicator is the relative strength indicator used to combined with comparison with different demographic location, which helps the result of volatility in prices and the

last indicator is the average true range which produce the result in range of price and price of the crops is varying.

### G. Pivot Point Indicator

Pivot points are the data indicators used to find the whole trend which is going on in the market for the different time intervals or frames. The pivot point can be defined as the average of all high, low and close prices of the prior day. This data indicator is mostly used by the people which are involved in the financial related market. This data indicator shows the support and resistance values. The pivot point for the data can be calculated using the formula below:

- Pivot Point = Yesterday's (high + low + close) / 3
- Support Price 1 = Pivot Point + (0.382 \* (max\_price - min\_price))
- Support Price 2 = Pivot Point + (0.618 \* (max\_price - min\_price))
- Support Price 3 = Pivot Point + (1 \* (max\_price - min\_price))
- Resistance Price 1 = Pivot Point - (0.382 \* (max\_price - min\_price))
- Resistance Price 2 = Pivot Point - (0.618 \* (max\_price - min\_price))
- Resistance Price 3 = Pivot Point - (1 \* (max\_price - min\_price))

### H. Relative Strength Indicator

In financial market the value or price greater than 70 is considered to buy more and if the value of price of less than 30 than it is considered to more sell. To understand the RSI, there are four main components are needed for the calculations of the RSI values, these are Relative Strength, Mean Gain, Mean loss and the RSI calculation is generally done for 14 periods days but in our project we taken it as 30 period of days which can't be in the negative value and as shown as positive values.

The calculation of the RSI is generally done using the below formula for the 1st period:

$$\text{1st Average Gain Price} = \text{Sum of all the Gains over the past 29 periods} / 30$$

$$\text{1st Average Loss Price} = \text{Sum of all the Losses over the past 29 periods} / 30$$

The calculation of the RSI is generally done using the below formula for the 2nd period:

$$\text{Average Gain} = [(\text{Prior Average Gain}) \times 29 + \text{Present Gain}] / 30$$

$$\text{Average Loss} = [(\text{Prior Average Loss}) \times 29 + \text{Present Loss}] / 30$$

$$\text{Strength} = \text{Average Gain Price} / \text{Average Loss Price}$$

$$\text{RSI} = 100 - (100 / (1 + \text{R.Strength}))$$

### I. Average True Range Indicator

The whole idea of candlesticks charting is price movement and price volatility associated with it. Average true range (ATR) finds the volatility difference and helps in deciding the volatility direction. This is just a range in the underlying trades. To make it more clearly, it's a price level is critical for the underlying next price movement and can generate huge volatility. Average true range for the data can be calculated by the difference between the high price and the low price. The average true range does not provide the price and its movement instead of the price provides the volatility in its price.

In general, the actual average range (actual average range, ATR) is calculated for 14 days of period and can be used for the calculation for the different intervals. The ATR calculation can done only on the daily base data.

$$\text{Current Atr} = ((\text{Previous Atr} * 13) + \text{Current TR}) / 14$$

**IV. EXPERIMENT RESULTS & ANALYSIS**

The final output of the system is able to give suggested price changes, price range, and level of volatility in prices and visual candlestick representation of prices. To find the degree of variance and volatility in prices in different demographic of India and also in different time frames. This data could really help farmers, small and medium enterprises in decision making, risk analysis, efficient returns on capital employed. Some interesting findings, how prices and indicators were generating very different signals at a common time at different demographics. Prices are very volatile in metropolitan cities where there is huge uncertainty of migrating population or seasoned states. This findings could help in pre-planning and improving food security by governments for monitored demographics. There are also regular patterns which have been found and could help take decisions. For example a classic rectangle pattern (prices move inside a range) was seen.

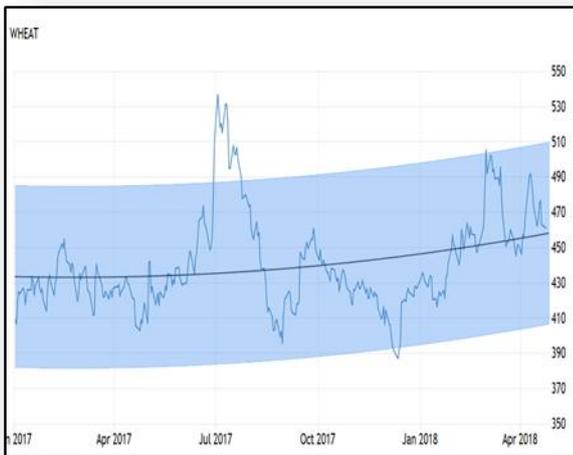


Figure 5: An Image of wheat price for one of the cities in last 4-6 months, and price trading in a rectangle pattern (range).

In Fig5, this would have been known that the prices will be moving in a range, spike in price above the range would have signalled a sell for the user every time when the prices touch the upper band of the pattern. Fig 6, showing the values of RSI values for the city along with the range and Fig 7 showing the prices of the wheat crop on the 30 days period shown using candle stick.



Figure 6: Visualization Chart showing the values of RSI values for the city along with the range.



Figure 7: Visualization chart for wheat crop prices of 30 days period shown using candle stick

**V. CONCLUSION & FUTURE WORK**

Able to implement an idea of building an application which can take input any type of commodity data available on government portal or different sources and help individuals to get better return on capital employed using candlestick charts, patterns, indicators, algorithms to better predict the accuracy of price change.



## REFERENCES

1. Hongbing Ouyang, (2011) The information content of high prices, low and fixed. This paper attempts to reveal the predictability of candle analysis at future prices and volatility. We find that high and low price series are more sensitive to price volatility than near prices. **IEEE Xplore, DOI:** 10.1109/AIMSEC.2011.6011140.
2. Donato Renato De, Delfina Malandrino, Giuseooina Palmieri, DatalEt-ecosystem provider (DEEP), an extensible and scalable Edge-centric architecture to display Open Data, retrieved in real-time from institutional open dataportals. **IEEE Xplore, DOI:** 10.1109/CeDEM.2017.14.
3. The Honorable Chiung-Leon Lee, Yi-Ching Liaw Hsu and Lindroos, Making investment decisions using candles and diffuse genetic algorithm pattern, **IEEE International Conference on Fuzzy System, 2011. DOI:** 10.1109/FUZZY.2011.6007707.
4. Hu Yang Shulin and Jieping, Research and Implementation of Web Services in Android Volley Network Communication. **IEEE Xplore, 2014 11th International Conference on Service Systems and Service, Management. 2014, DOI:** 10.1109/ICSSSM.2014.6943373.
5. Leo Rizky Julian and Friska Natalia, 2015 using computer parts scraping web mount comparator.
6. Kim T, Kim HY (2019) Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data. **PLoS ONE 14(2):** e0212320. <https://doi.org/10.1371/journal.pone.0212320>



**Dr.M.S.Abirami**, received an M.Tech in Computer Science from SRM University, Chennai, India, an MCA from Bharathidasan University, Trichy and completed her Ph.D. in the Image Processing at Bharathiyar University, Coimbatore, India. She has 19+ years of teaching experience. Currently, she is an assistant professor (Senior Grade) in the Department of Software Engineering, SRM Institute of Science and Technology, India. She is the author of 10 international journal papers and 12 conference papers. Her research interests include Image Processing, Parallel and Distributed Computing, Data Mining and Machine Learning.



**Dr.S.Metilda Florence**, completed PhD in Video Processing area from Bharathiar University, Coimbatore, India. Received M.Tech degree in Computer Science from SRM University, Chennai, India and MCA from Bharathidasan University, Trichy, India. She has 18+ years of teaching and research experience. Currently she is an Assistant Professor (Senior Grade) in Information Technology Department at SRM Institute of Science and Technology, India. She is the author of 10 International journals and 8 International conference papers. Her research interests include Image processing, Data mining, Recommendation system, Machine learning and deep learning.

## AUTHORS PROFILE



**Dr.Uma Mohan**, received an M.Tech in Computer Science from SRM University, Chennai, India, an MCA from Bharathidasan University, Trichy and is completed her Ph.D in the area of Brain Computer Interface from Bharathiyar University, Coimbatore, India. She has 18+ years of teaching and research experience. Currently, she is an assistant professor (Selection Grade) in Department of Software Engineering, SRM Institute of Science and Technology, India. She is the author of 18 journal publications and 17 national and international conferences. Her research interests include brain computer interface, P300, Data Mining, Personalization, Recommendation System, Machine learning, Web services.