

Linear Regression Model For Stock Market Analysis and Prediction System

Samreen Fayaz, Vasudhan Reddy, Brijesh Kumar Patel

Assistant Professor, B.Tech student, B.Tech student

School of computer science and Engineering

Lovely Professional University, Phagwara, Punjab, India

Samreen.25435@lpu.co.in¹, reddyvasudhan@gmail.com², brijeshpatel4547@gmail.com³

Abstract

In this paper we are envisaging the stock Market by considering various aspects that are tangled in the forecast – bodily factors vs. Psychological, normal and illogical behavior, etc. All these aspects associate to create share prices instable and actual challenging to forecast with a great degree of precision. Our objective is to build the software/website which analyses former stock records of assured companies based on several weather factors affecting stock assessment. The website is intended to give the predictions on the stock prices of selected companies based on the weather or climatic conditions in the city/state. The stock prices are reflected as per the Bombay Stock Exchange (BSE). These predictions are based on the behavior of the stock price of the company on each day with respect to the climatic conditions on the respective day. The data used to predict the prices is of the recent year i.e., 2019 and 2018. We are using the ML Algorithms to generate the Regression model which will be used for the predictions and this model will be integrated on the web using the FLASK. Using the SPSS software for pre-analysis of the data to verify any outliers and also for generating the charts and graphs. Using Apache Spark API's for managing the data and manipulating it as per requirements.

Key-Words: ML (Machine Learning), MLLIB (Machine Learning Library of Apache Spark), DF (Data Frame), FLASK (Micro Web Framework), SPSS (A Statistical tool to find insights of the Data, By IBM), LR (Linear Regression a ML model).

Introduction

The project uses various languages and algorithms to manipulate and manage the data and make predictions for future based on the past data. We start by collecting the data related to stock market and prices and weather data for the Dates in the stock prices dataset[1]. We merge the stock data and weather data. We then create a regression model i.e.,

LR-model which will be used for predicting the future. We divide our data into 2 parts i.e., train and test data. We train our model with the train data we have and then test the data with the remaining part of the dataset. Once the model is created we have to implement the model in web using the FLASK and create a web API for the same. Then we take the input from the user which consists of the factors that were used for creating the model and we generate the prediction for the input provided by the user whether it is profitable to Invest or buy the stocks of the company or not.

- The project predicts and analyzes stock market on the basis of the weather or climatic factors.
- The analysis and predictions are confined to only the particular region i. e, Mumbai.
- According to the weather conditions in the region, stock market analysis is done based on the past data of cost of the stocks and the future predictions are drawn based on the analysis, this is done by **ML** models.

The linear regression model uses the correlation factor to draw insights and to predict the future [2,3].

- i. If Correlation (r) is positive that means the variables are moving in same direction i.e., if one increases other also increases.
- ii. If Correlation (r) is negative that means the variables are moving in Opposite direction i.e., if one increases other decreases.
- iii. If Correlation (r) is Zero then there is no relationship between the variables.

The below image shows the correlation values and their meanings for interpreting the results :

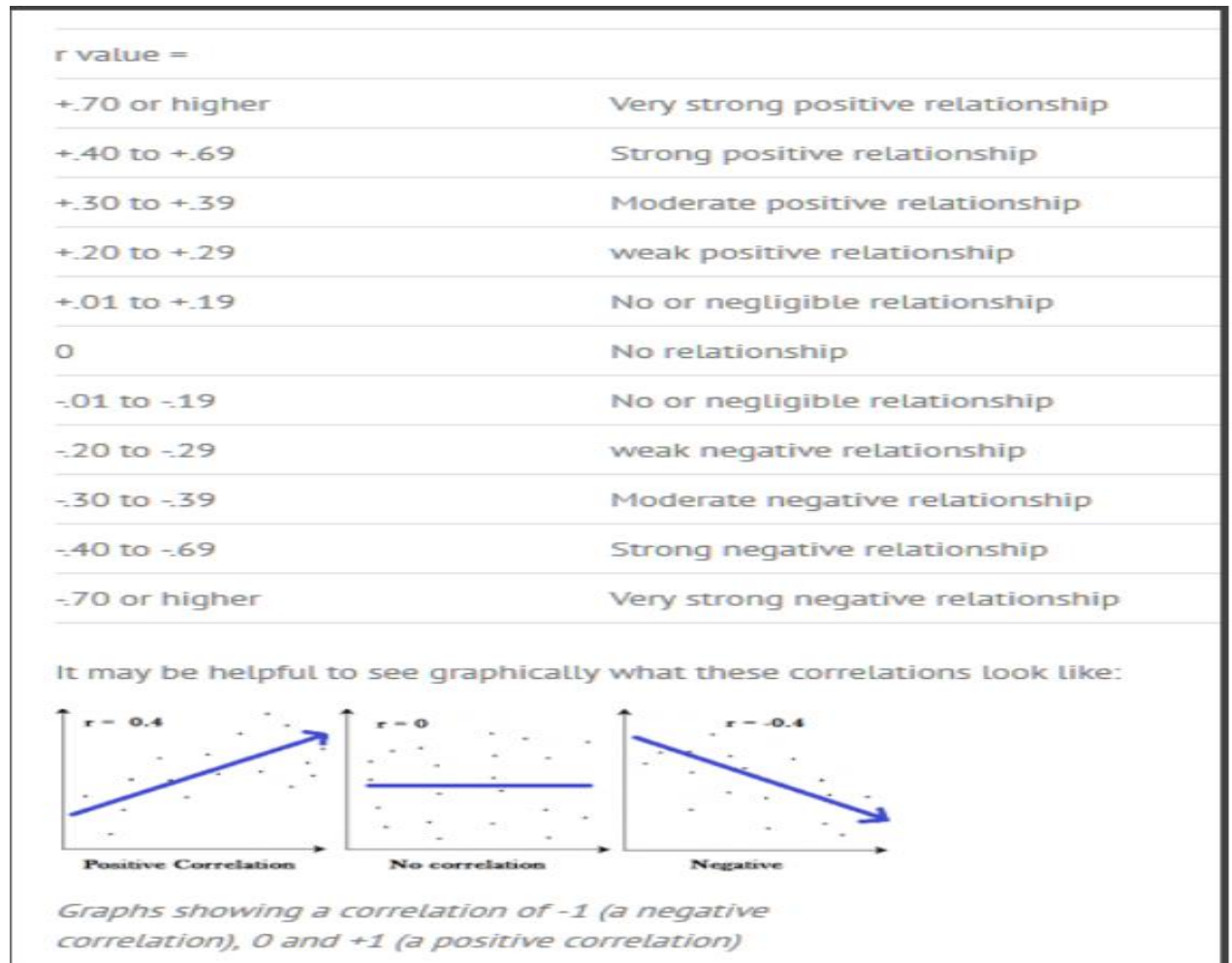


Figure 1 Types of correlations

Literature Survey

Stock and weather prediction, is financial markets influenced by the weather? Are they correlated with each other? Fix stock prices and equities dealing transaction capacities-track a sketch in relation to diverse weather situations? If there is a correlation, what could be driving it? This investigation seeks to achieve some vision interested in this area. Beyond the evident influence of great weather it besides requests if there is normalcy from one day to other, variations that occur from one to other season and instabilities have certain type of impact on behavior of the stock market. Lot of tools and trends are there that

can do analysis of trends and statistical on the stocks. But we could find only a single tool that could investigate weather data and the stock data.

The investigation community's judgments on the area of interactive finance appear to be shared. Certain say that there absolutely do not have any correlation between weather and several researchers say there exists. With it as the research gap where we can explore and find something new. It's relevant to exclude maintain tornadoes and tempests as these occurrences shake the market harshly with their impression. The information for the historical stock data used here is obtained as of Yahoo Finance.

The tool is created using the R-Language and is just an analyzer tool, which is just for analyzing weather there is any relation between the weather and stock prices. The API which we propose is used for the future predictions also.

Proposed System

According to our proposed methodology, the accuracy and precision of the system is mostly based on the data, more the data more is accuracy of the model and more is number of correct predictions drawn. More over the accuracy also depends on how the data is distributed, any missing values in the data, any outliers, etc. The system works as follows:

1. Input is taken from the users , which consists of certain weather factors such as Average Temperature , Humidity, Total Rainfall, Wind speed, Average Visibility
2. Once the user inputs the values for the above variables, the predictions are drawn considering all the inputs entered.
3. The output will be displayed on the monitor, if the investment would be profitable or not profitable considering the weather factors.
4. User can also view the analysis of the past data.

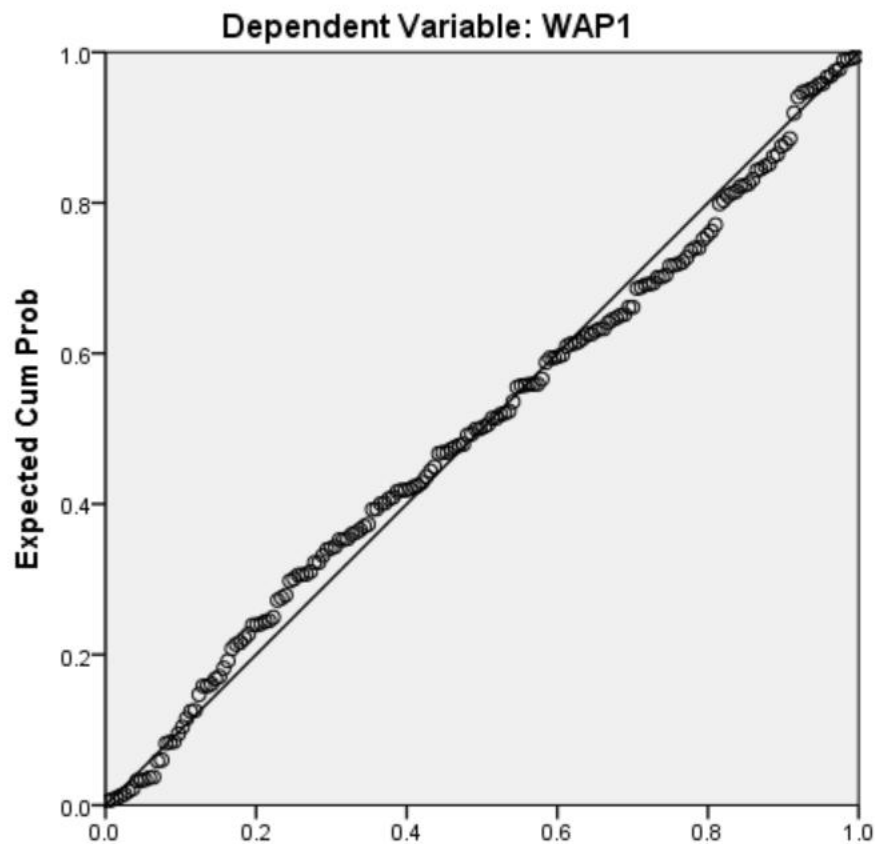
We are using the linear regression model for carrying out the analysis and predicting the future, using the FLASK we have been deploying the model on the web so that it could

be used as a Web API. We can build the website using html, CSS, JavaScript, PHP etc. The website is connected to a database to maintain User log data and other details.

Features Used for Predictions:

The features that we have selected for investigation are as follows:

- Average Temperature
- Humidity
- Total Rainfall
- Wind speed
- Average Visibility



Plot showing the linear relationship between dependent and independent variable

Analysis Results

a) Dependent and Independent Variables:

Tables below are showing the dependent and independent variables and the linear relationship values.

Table1,2

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	WindSpeed, TotRain, Tavg, Avgvis, Humidity ^b	.	Enter

a. Dependent Variable: WAP1

b. All requested variables entered.

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.631 ^a	.398	.381	22.96689674

a. Predictors: (Constant), WindSpeed, TotRain, Tavg, Avgvis, Humidity

b. Dependent Variable: WAP1

Predictors => Independent

Dependent => Avg Stock Price

b) Analysis :

Table 3,4 shows fitness value of the model and derives the equation of the regression line.

Table3,4

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	61502.102	5	12300.420	23.319	.000 ^b
	Residual	92836.189	176	527.478		
	Total	154338.291	181			

a. Dependent Variable: WAP1

b. Predictors: (Constant), WindSpeed, TotRain, Tavg, Avgvis, Humidity

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	116.393	24.347		4.781	.000
	Tavg	2.434	1.061	.167	2.294	.023
	Humidity	-.608	.152	-.351	-4.010	.000
	TotRain	.010	.051	.014	.206	.837
	Avgvis	11.565	2.669	.314	4.332	.000
	WindSpeed	-1.236	.603	-.158	-2.048	.042

a. Dependent Variable: WAP1

Dataset View

```
1 df.printSchema()
```

```
root
|-- Date: string (nullable = true)
|-- Open Price: double (nullable = true)
|-- High Price: double (nullable = true)
|-- Low Price: double (nullable = true)
|-- Close Price: double (nullable = true)
|-- WAP: double (nullable = true)
|-- Tavg(T): double (nullable = true)
|-- Humidity(H): integer (nullable = true)
|-- Total_Rainfall(PP): double (nullable = true)
|-- Avg_Visibilty(VV): double (nullable = true)
|-- WindSpeed(V): double (nullable = true)
```

Conclusion

From the above analysis we have concluded that $R=0.63$ which is the relationship value between dependent and independent variables i.e., positive and it's between 0.4- 0.7, Hence the correlation between the features and the stock price is Strong and Hence we can say that there is a relationship between the weather factors and the stock prices. Hence we conclude our project by drawing the predictions based on the features provided. Stock Market Analysis of stocks using data mining will be useful for new investors to invest in stock market based on the various factors considered by the software. Stock market includes daily activities like Sensex calculation, exchange of shares. The exchange provides an efficient and transparent market for trading in equity, debt instruments and derivatives. Our software will be analyzing Sensex based on company's stock value of the particular day. The stock value of company depends on many factors but we are only concerned with the weather condition of the particular day in particular region. After the analysis is created, then using the analysis result we will be predicting the stock market values at the real time which may be of huge help for the stock market investors. As stock market is a worldwide network and the analysis is created for the particular region, the predictions may or may not be applicable for everyone. But this may give an idea to the investors that how the stock price varies based on weather conditions.

References

- [1] Dr. S. Santhosh Baboo and I. Kadar Shereef, "An Efficient Weather Forecasting System using Artificial Neural Network", International Journal of Environmental Science and Development, Vol. 1, No. 4, October 2010.
- [2] Arvind Sharma, Prof. Manish Manoria, "A Weather Forecasting System using concept of Soft Computing," pp.12-20, 2006

- [3] N. Sharma, J. Gummeson, D. Irwin, and P. Shenoy, "Leveraging Weather Forecasts in Energy Harvesting Systems," University of Massachusetts Amherst, Tech. Rep., September 2011.
- [4] Mohsen Hayati, and Zahra Mohebi," Application of Artificial Neural Networks for Temperature Forecasting," World Academy of Science, Engineering and Technology 28 2007.
- [5]Navin Sharma, Pranshu Sharma, David Irwin, and PrashantShenoy,"Predicting Solar Generation from Weather Forecasts Using Machine Learning",virtual power plants,distributedgenerations,microrids,renewables and storage(IEEE SmartGrid-Comm),2011
- [6]J. Dahya ,A.A. Lonie ,D.M. Power,"The Case for Separating the Roles of Chairman and CEO: An Analysis of Stock Market and Accounting Data", Oct 2006
- [7]George J. Feeney,Donald D. Hester," Stock Market Indices: A Principal Components Analysis",1964
- [8]Dr. Kunwar Singh Vaisl,Dr. Ashutosh Kumar Bhat,"An Analysis of the Performance of Artificial Neural Network Technique for Stock Market Forecasting",May 2014
- [9]Kalyani Dacha, "Causal Modeling of Stock Market Prices using Neural Networks and Multiple Regression: A Comparison Report" Finance India, Vol. xxi , No.3,p-923-930,2007
- [10]National Stock Exchange (NSE), www.nse-india.com