

---

# **STOCK MARKET ANALYSIS USING CORRELATION AND CLUSTERING ALGORITHMS**

---

April 27, 2016

Submitted by :

Aman Bakshi (Roll No:13)

Jallipalli Vinay(Roll No:14)

Department of Information and communication technology  
MANIPAL INSTITUTE OF TECHNOLOGY

# Contents

Abstract . . . . .	2
Introduction . . . . .	2
Domain Introduction . . . . .	2
Functionality and design . . . . .	3
Tools used . . . . .	3
Implementation . . . . .	4
Conclusion . . . . .	5
References . . . . .	5

---

## ABSTRACT

In this paper we present a novel data mining approach to predict behavior of stock market trends. In our paper, the trends and patterns in stock market are shown and analyzed, on the basis of the current and old stock market data, available. The patterns are evaluated using clustering algorithm and correlation analysis. Some patterns and trends always emerge in stock markets, which affect prices of all related stocks. By using the cluster information and correlation analysis, our approach predict the stock trends and patterns effectively in the real world market.

## INTRODUCTION

The stock market is an ever-changing space with highly irregular changes in patterns and its behavior. Analyzing stock market behavior is an important economic need. The process applied here, in analyzing stock market consists of clustering and searching for any occurring anomalies. This will help in easy understanding of stock market and predicting and evaluating the patterns occurring in the stock market. Subsequently, leading to a better understanding how the stock market works and how the stock prices changes

## DOMAIN INTRODUCTION

The clustering algorithm is consists of **data mining** and **cluster creation** These two terms are inherently related. The data mining part, consists of extracting the data in a useful format and cluster creation, consists of dividing the data into interesting clusters.

---

## FUNCTIONALITY AND DESIGN

The process of clustering can be divided into following steps:

1. **Data acquisition** : The data used in the algorithm, is the stock sheet, consisting of one year's worth of stock values of a company. The data used has following categories:
  - a) Date : Date corresponding the stock values.
  - b) Open : The price of the stock value at which it opened at the corresponding date.
  - c) High : The highest value achieved by the stock at the corresponding date.
  - d) Low : The lowest value achieved by the stock at the corresponding date.
  - e) Close : The price of the stock at which it closed at the corresponding value.
  - f) Volume : The number of the stocks exchanged.
  
2. **Cluster formation** : Once the data is extracted, it is divided into clusters, by implementing K-means, on the stock values. The clusters created are thus, the result of the implementation

## TOOLS USED

We have primarily used **Python** as our base programming language. We have used various additional modules of python such as **math** and **scipy**.

---

## IMPLEMENTATION

We first fetch the data from data source then make a pre processing to the data, after we feed it into the k-means algorithm to find anomalies. We can then make predictions according to the position of the anomalies and evaluate the result.

### 1. Data preprocessing:

Pre-processing in data mining played essential role for enhancing data quality. The basic concept behind is that, learning with accurate and high quality data may provide more efficient classification results as compared to learning with poor quality of data.

Data mining is an approach to find the meaningful patterns from data. This meaningful content may helpful for decision making, classification and large scale data analysis. In data mining the main and basic element is data. Mining of data and information recovery is directly depends upon data. Therefore, learning process of a data mining algorithm is majorly depends upon the type of data and quality of data.

we are reading the data from **CSV** file which initially contains only **DATE, OPEN PRICE, CLOSE PRICE, VOLUME**. we process the data to get change in open and close price for all the entries in the data file. we used the python libraries to read the data from CSV file.

### 2. Algorithm Implementation:

The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume  $k$  clusters). The main idea is to define  $k$  centers, one for each cluster. These centers should be placed in a proper way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate  $k$  new centroids as barycenter of the clusters resulting from the previous step. After we have these  $k$  new centroids, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop we may notice that the  $k$  centers change their location step by step until no more changes are done or in other words centers do not move any more.

---

## CONCLUSION

we found anomalies in the data by clustering the data using k means algorithm. after clustering the points which are far to clusters are known as outliers.

## REFERENCES

### Base paper:

<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=7310722>

### Plotting graphs in python using libraries:

<http://matplotlib.org/examples/index.html>

[http://matplotlib.org/examples/shapes\\_and\\_collections/scatter\\_demo.html](http://matplotlib.org/examples/shapes_and_collections/scatter_demo.html)

### For date and time calculation:

[http://www.tutorialspoint.com/python/time\\_mktime.htm](http://www.tutorialspoint.com/python/time_mktime.htm)

<https://docs.python.org/2/library/datetime.html>

[http://www.tutorialspoint.com/python/time\\_strptime.htm](http://www.tutorialspoint.com/python/time_strptime.htm)

### For documentation using Latex:

We also used online tutorials to look up for LaTeX.

And Document was made online using

<https://www.overleaf.com/dash>

We used various online tutorials for learning pyhton syntax.