JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Stock movement analysis using Clustering Techniques

[1]GANESH ROHAN N, [2]DHIWAKAR JAYAPRAKASH KUSUMA

[1] Master of Technology – Artificial Intelligence, [2] Master of Science – Artificial Intelligence
[1 2] REVA Academy for Corporate Excellence (RACE),
[1 2] REVA University, Bengaluru, India

Abstract : While stocks get clustered based on Industry, a technical analyst will look for pattern of movements in stocks across industry. This paper will show an example of how this cluster analysis can be done, based on some out of the box features, and some engineered features. We will also delve in to comparing the out of the box clustering algorithms like K-Means vs customized algorithms for clustering to see how the clustering works. Using these techniques one can identify groups of stocks which highlight movement patterns, and take investment decisions e.g. You can consider investing in stocks based on the movement of the other stocks in the same cluster.

***Index Terms* – Clustering, K-Means, Customer Algorithm, Stock Movement**

I. INTRODUCTION

Investment strategies in secondary market or share markets come in different varieties. At a prominent level, these strategies get classified under Fundamental analysis and Technical Analysis. Under Technical analysis, stock movements are predicted using varied features and one suck is clustering technique. A stock gets in to a cluster based on the movements that it has exhibited in the past, and clustered with the other stocks that has exhibited similar movements. The analysis itself is based on variety of features (both out of the box, and engineered).

II. DATA METHODOLOGY

2.1 Data Collection

For this study, stock data has been downloaded via Yahoo Finance module for the past 2 years. Have chosen random stocks from diverse industry verticals to check and observe if they can be clustered – e.g., from Information Technology vertical we have chosen companies like Apple, Microsoft etc. and from Banking vertical we have chosen companies like Bank of America, Wells Fargo etc.,

2.2 Data Processing and Feature Engineering

Based on the daily data that has been downloaded through Yahoo Finance, following features have been engineered. We can certainly add more features and perform a PCA to determine the apt feature that work for the study – but, this is beyond the scope of this study.

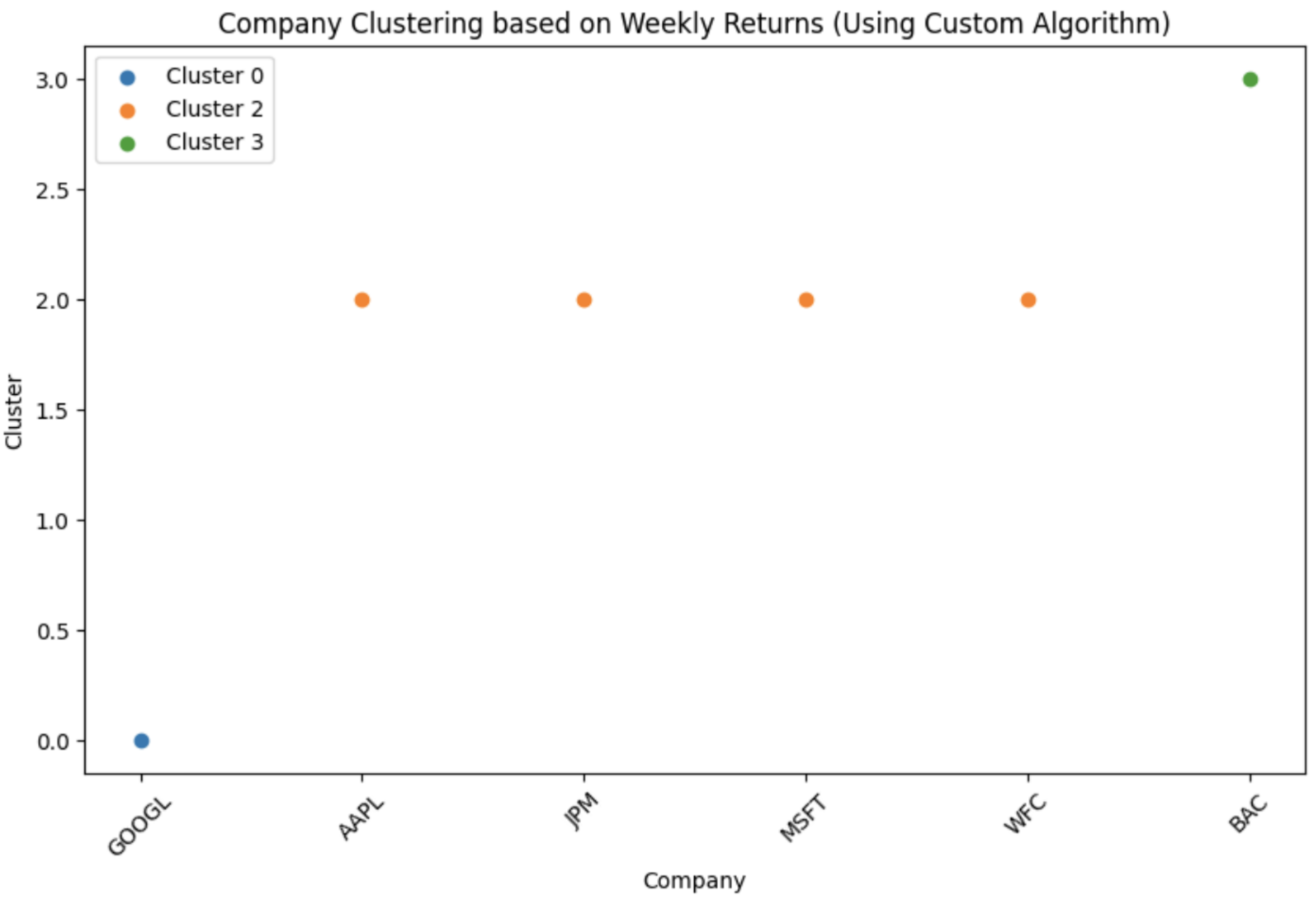
- Weekly Open: Opening price for that week
- Weekly Close: Closing price for that week
- Weekly Price Movement

III. MODEL METHODOLOGY

3.1 Results of Descriptive Statics of Study Variables

With the engineered features, we applied the ready to consume K-Means algorithm and results are shown in the figure below. Based on K-Means, we managed to group the companies in to 4 clusters like shown below (number of clusters is configurable). One can already observe some interesting facts like how a banking company like Wells Fargo is clustered with Information Technology companies like Apple and Microsoft, and how JP Morgan and Bank of America is not moving in tandem with their peers in the same industry etc.,

As the next step a custom clustering algorithm was created – The code defines thresholds to categorize the mean weekly returns into 4 clusters based on their distance from the overall mean, using the standard deviation as a measure of dispersion. The thresholds are adjusted to create four distinct clusters. The results can be observed below



REFERENCES

Basalto, Nicolas & Carlo, Francesco. (2006). Clustering financial time series. 10.1007/4-431-28915-1_46.

T. Wahyuningrum, S. Khomsah, S. Suyanto, S. Meliana, P. E. Yunanto and W. F. Al Maki, "Improving Clustering Method Performance Using K-Means, Mini Batch K-Means, BIRCH and Spectral," 2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 2021, pp. 206-210, doi: 10.1109/ISRITI54043.2021.9702823. keywords: {Seminars;Error analysis;Clustering methods;Training data;Pressing;Iterative methods;Intelligent systems;clustering;KNN;K-Means;BIRCH;Spectral},

S. Bhatia, "New improved technique for initial cluster centers of K means clustering using Genetic Algorithm," International Conference for Convergence for Technology-2014, Pune, India, 2014, pp. 1-4, doi: 10.1109/I2CT.2014.7092112. keywords: {Clustering algorithms;Biological cells;Genetic algorithms;Sociology;Statistics;Algorithm design and analysis;Complexity theory;Clustering;Data Mining;K means clustering algorithm;Initial cluster centers;Genetic Algorithm;Outliers;Centroid},

M. Mardi and M. R. Keyvanpour, "GBKM: A New Genetic Based K-Means Clustering Algorithm," 2021 7th International Conference on Web Research (ICWR), Tehran, Iran, 2021, pp. 222-226, doi: 10.1109/ICWR51868.2021.9443113. keywords: {Machine learning algorithms;Clustering methods;Software algorithms;Clustering algorithms;Stochastic processes;Web mining;Partitioning algorithms;Machine Learning;Clustering;K-Means Algorithm;Genetic Algorithm;Data Mining},

K. P. Sinaga and M. -S. Yang, "Unsupervised K-Means Clustering Algorithm," in IEEE Access, vol. 8, pp. 80716-80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
keywords: {Clustering algorithms;Indexes;Linear programming;Entropy;Clustering methods;Unsupervised learning;Machine learning algorithms;Clustering;K-means;number of clusters;initializations;unsupervised learning schema;Unsupervised k-means (U-k-means)},

J. Jeyachidra, T. Logesh, K. Nandhini and R. Krithiga, "Hybrid K-Means Clustering for Training Special Children using Utility Pattern Mining," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ICECONF57129.2023.10083709. keywords: {Training;Performance evaluation;Clustering methods;Clustering algorithms;Sensitivity and specificity;Knowledge discovery;Proposals;K-Means;Clustering;Grouping;Data Mining;Genetic Algorithm;Hybrid K-means;Accuracy},