

Introduction

Instructions

Select an Industry

Module 1

Module 2

Module 3

Module 4

Module 5

Module 6

Interview

Certificate

My Jupyter Notebook

Module 6

Clustering for Diversification analysis

Clustering is a method of unsupervised learning and is a common technique for statistical data analysis used in many fields.

Clustering is a Machine Learning technique that involves the grouping of data points. Given a set of data points, we can use a clustering algorithm to classify each data point into a specific group. In theory, data points that are in the same group should have similar properties and/or features, while data points in different groups should have highly dissimilar properties and/or features.

In financial Markets, Cluster analysis is a technique used to group sets of objects that share similar characteristics. It is common in statistics, but investors will use the approach to build a diversified portfolio. Stocks that exhibit high correlations in returns fall into one basket, those slightly less correlated in another, and so on, until each stock is placed into a category.

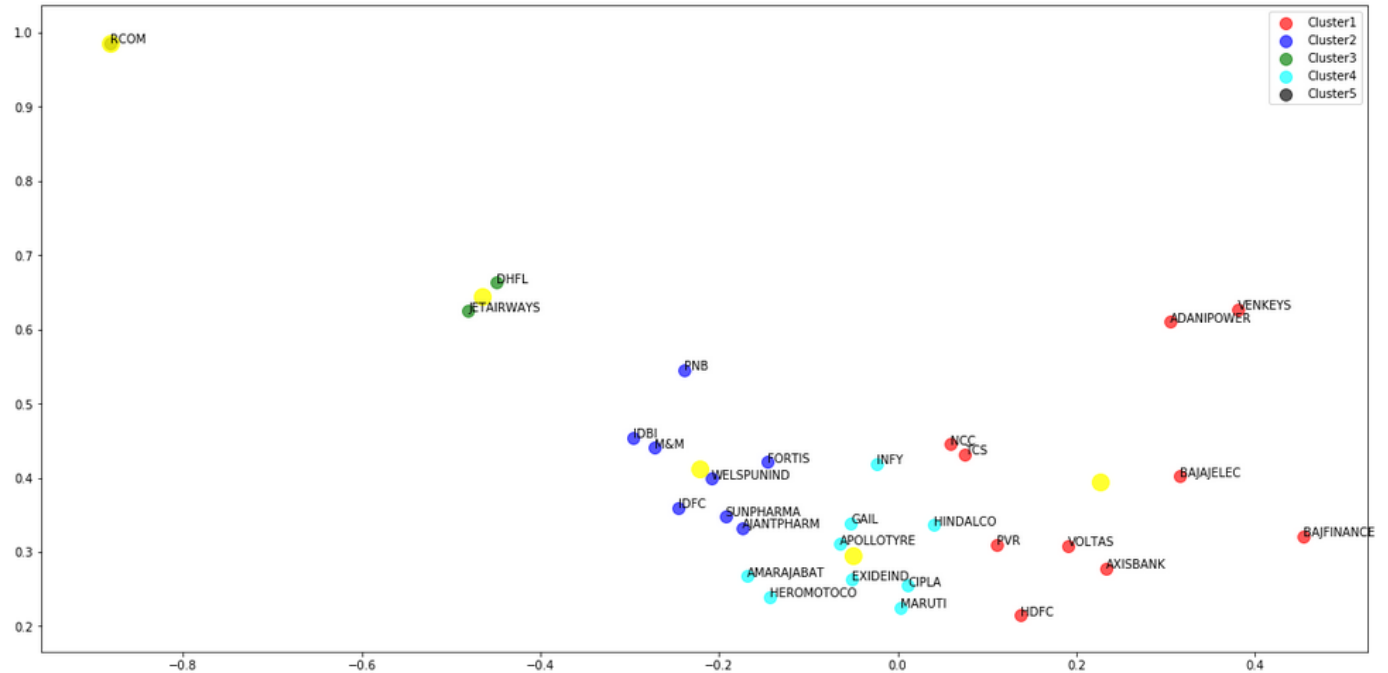
Problem Statements

6.1 Create a table/data frame with the closing prices of 30 different stocks, with 10 from each of the caps

6.2 Calculate average annual percentage return and volatility of all 30 stocks over a theoretical one year period

6.3 Cluster the 30 stocks according to their mean annual Volatilities and Returns using K-means clustering. Identify the optimum number of clusters using the Elbow curve method

6.4 Prepare a separate Data frame to show which stocks belong to the same cluster



Before you proceed, please take this **quick quiz** on Module five and six for us to assess your learning

Getting stuck and need a little guidance? Get access to the **guidance program** (<https://careerlauncher.com/cl-online/ProductDesc.jsp?prodeid=0K4XsvNcqbM%3D&prodCat=DATA%20ANALYTICS&prodGroup=Online%20Coaching>) from our machine learning faculty!