

MACHINE LEARNING EN FINANZAS

Problem Set 4

INTRODUCCIÓN

Dentro de una hora tienes un comité de inversiones en el que discutirán inversiones en FCIs. Tu tarea es llevar al comité un listado de fondos clasificados según tipo. Es la primera vez que te asignan esta tarea por lo que sólo dispones de un reporte que te llega diariamente con 58 fondos que figuran clasificados como Money Market (MM), Renta Fija (RF) y Equity (EQ).

Dispones de poco tiempo para evaluar si la clasificación es correcta o no, por lo que decides hacer caso omiso al reporte y utilizar herramientas de clustering para sacar conclusiones rápidas.

El reporte contiene además del tipo de FCIs, datos de volatilidad para el último mes, y retornos para los últimos 3 y 6 meses. Estas tres variables son lo que utilizarás para el clustering.

REQUISITOS

Implementar un script para correr interactivamente que:

1. Cargue los datos del dataset 'fondos_comunes.csv' en un `pandas.DataFrame`
2. Realice preprocesamiento de los datos de ser necesario. Querés estar bien cubierto ante preguntas que puedan comprometer la metodología que usaste.
3. Decides utilizar 'K-Means' y 'Agglomerative Hierarchical Clustering' como modelos, y evaluarás la cantidad correcta de clusters a usar dentro de un rango de 2 a 10 utilizando el 'Calinski-Harabasz Score'. Para ello imprimirás los dos gráficos correspondientes a cada modelo. Usarás un loop para no desperdiciar tiempo copiando y pegando con potenciales errores.
4. Luego de interpretar los gráficos y establecer la cantidad de clusters con la que te sentís cómodo, construyes un listado comparativo entre la clasificación original y lo que tus modelos han hallado. Cualquier cosa rara que encuentres prefieres dejarla fuera de las propuestas.

Listo para el comité y cualquier pregunta sobre el tema!

ÉXITOS!