

# PROJECT “HOSE COMPANY SEGMENTATION” - Insight

## I/ VỀ CÁC CỤM/NHÓM CÔNG TY

Ta có thể đưa ra đặc điểm các Cụm (Nhóm) Công ty sàn HOSE giai đoạn 01/2023 - 09/2023 như sau:

- Cụm 0:
  - Tăng trưởng nhiều về mặt doanh thu và cả tổng số nợ. Tuy nhiên mức lợi nhuận chưa có sự gia tăng tương xứng.
  - Các chỉ số tăng trưởng/suy giảm với biên độ lớn và không rõ ràng.
  - Đánh giá thị trường khá tích cực thể hiện qua chỉ số PE đa số là tăng dần. Tuy nhiên EPS đa số giảm đều phản ánh thực tế Các Cty nhóm này chưa hoạt động hiệu quả.

Nhận xét: Cụm này đặc trưng chính là tính **Biến Động**, thể hiện trọng tâm là đầu tư phát triển dù có phụ thuộc vào vay mượn bên ngoài và chấp nhận lợi nhuận thấp/tăng trưởng chậm. Đây là điểm thường thấy ở những **Công ty non trẻ mới tham gia thị trường**.

Đề xuất: Cần nhắc theo dõi thêm trong 1 thời gian vừa phải trước khi tiến hành cho vay/đầu tư trái phiếu/cổ phiếu ...

- Cụm 1:
  - Không có sự tăng trưởng nào đáng kể về doanh thu hay lợi nhuận.
  - Về mặt đánh giá thị trường không mấy tích cực, phản ánh qua EPS giảm đều theo thời gian và PE không có thay đổi nhiều.
  - Tuy nhiên điểm sáng Cụm này là tổng số nợ giảm dần theo thời gian.

Nhận xét: Cụm này đặc trưng chính là sự **Ổn Định**, thể hiện trọng tâm là sử dụng hiệu quả nguồn vốn tự có và giảm sự phụ thuộc vào vay mượn bên ngoài, dù có thể ảnh hưởng lợi nhuận. Đây là điểm thường thấy ở những **Công ty lâu đời dày dạn kinh nghiệm**.

Đề xuất: Phù hợp với việc cho vay thông thường.

- Cụm 2:
  - Tăng trưởng đều về doanh thu và lợi nhuận song song với giảm dần tổng số nợ theo thời gian.
  - Đánh giá thị trường nói chung tích cực, phản ánh qua EPS và PE tăng dần qua thời gian.

Nhận xét: Cụm này đặc trưng chính là sự **Phát Triển**, thể hiện trọng tâm là sử dụng hiệu quả nguồn vốn tự có song song giảm sự phụ thuộc vào vay mượn bên ngoài. Đây là điểm thường thấy ở những **Công ty trưởng thành đang chiếm lĩnh thị trường**.

Đề xuất: Đây là Nhóm cần chú ý nhất, phù hợp với việc cho vay thông thường lẫn đầu tư trái phiếu/cổ phiếu hay dạng khác.

## II/ VỀ SCRAPING

1. Với web static: data dạng bảng > get\_html(), data khác thì dùng BS4.
2. Với web dynamic: dùng Selenium.
3. Trong khi scrape chú ý nghỉ gián đoạn với sleep().

## III/ VỀ THUẬT TOÁN

1. Với data phân bố rõ ràng > Kmean & Hierarchical: dễ thực hiện & nhanh.
2. Với data có nhiễu > DBSCAN: nhận diện được nhiễu & loại bỏ nhiễu.
3. Nên sử dụng ít nhất 3 Model & 3 Metric để so sánh và chọn ra K tối ưu. Sau đó mới tiến hành phân cụm và gán label.