

TRƯỜNG ĐẠI HỌC KINH TẾ TP. HCM
KHOA CÔNG NGHỆ THÔNG TIN KINH DOANH

-----□□□□-----



TIỂU LUẬN
MÔN KHOA HỌC DỮ LIỆU

ĐỀ TÀI:
NGHIÊN CỨU VỀ BÁO CÁO TÀI CHÍNH CỦA
CÁC CÔNG TY HOA KỲ

Giảng viên: Th.s Nguyễn Mạnh Tuấn

Lớp học phần: 22C1INF50905965

Sinh viên thực hiện:

Họ và tên	Mã số sinh viên
Lê Thị Huyền	31211021802
Huỳnh Thị Thanh Ngân	31211025108
Lê Trần Quỳnh Như	31211021923
Trần Hoàng Xuân Như	31211021929
Phạm Hồng Quyên	31211023949

Thành phố Hồ Chí Minh, ngày 9 tháng 10 năm 2022.

DANH MỤC HÌNH ẢNH

Hình 1: Thông tin dữ liệu đầu vào	11
Hình 2: Mô hình tiền xử lý dữ liệu bằng Orange	12
Hình 3: Mô hình phân tách dữ liệu	13
Hình 4: Mô hình bài toán 1	15
Hình 5: Mô hình xây dựng bài toán 2	20
Hình 6: Mô hình xây dựng bài toán 3	24
Hình 7: Kết quả phân cụm theo phương pháp Hierachial Clustering	25
Hình 8: Chỉ số Silhouette của phương pháp Hierachical Clustering	25
Hình 9: Kết quả phân cụm theo phương pháp K-Means	26
Hình 10: Chỉ số Silhouette của phương pháp K-Means	27

DANH MỤC BẢNG

Bảng 1: Mô tả dữ liệu bảng	9
Bảng 2: Kết quả Confusion Matrix của bài toán 1	17
Bảng 3: Kết quả Test & Score của bài toán 1	18
Bảng 4: Kết quả dự báo của bài toán 1	19
Bảng 5: Kết quả Confusion Matrix của bài toán 2	21
Bảng 6: Kết quả Test & Score của bài toán 2	22
Bảng 7: Kết quả dự báo của bài toán 2	23
Bảng 8: Bảng kết quả phân cụm của bài toán 3	28

MỤC LỤC

CHƯƠNG I: TỔNG QUAN	4
1. Lời mở đầu	4
2. Giới thiệu đề tài	4
3. Mục tiêu nghiên cứu	5
3.1. Mục tiêu tổng quát	5
3.2. Mục tiêu cụ thể	5
4. Đối tượng nghiên cứu	5
5. Phương pháp nghiên cứu	5
5.1. Phương pháp phân lớp	5
5.2. Phương pháp phân cụm (Clustering)	7
6. Mô tả dữ liệu	7
CHƯƠNG II: QUY TRÌNH THỰC HIỆN VÀ KẾT QUẢ	11
1. Tiền xử lý	11
1.1. Tổng quan:	11
1.2. Quy trình thực hiện	11
<i>Hình 1. Thông tin dữ liệu đầu vào</i>	11
2. Xử lý các bài toán	12
2.1. Bài toán Phân lớp dữ liệu	12
2.1.1. Các phương pháp dự đoán và quy trình cụ thể	12
2.1.2. Xử lý các bài toán phân lớp	13
2.2. Bài toán phân cụm	23
2.2.1. Qui trình cụ thể	23
CHƯƠNG III: KẾT LUẬN	29
1. Kết luận	29
2. Những hạn chế	30
3. Hướng khắc phục	30

CHƯƠNG I: TỔNG QUAN

Đánh giá mức độ liên quan đến chuyên ngành: 1, Liên quan, vì nhóm sử dụng kiến thức chuyên ngành kế toán trong việc đánh giá báo cáo tài chính của hơn 200 công ty hàng đầu ở Hoa Kỳ. Từ đó giúp các nhà đầu tư có một cái nhìn tổng quan về tình hình tài chính, phát triển của công ty trong tương lai để đưa ra các quyết định đầu tư đúng đắn.

1. Lời mở đầu

“Báo cáo tài chính” là một cụm từ vô cùng quen thuộc đối với các nhà đầu tư, các doanh nghiệp, đặc biệt là với các nhân viên làm trong lĩnh vực kế toán, kiểm toán, tài chính. Nó “đại diện” cho cả một doanh nghiệp trong việc thu hút dòng tiền từ các nhà đầu tư. Có thể nói doanh nghiệp có một bảng báo cáo tài chính “đẹp” là một doanh nghiệp đang kinh doanh rất tốt trên lĩnh vực của mình. Tuy nhiên, việc các chỉ số đều tốt ví như lợi nhuận cao chưa chắc đã là một sự lựa chọn hoàn hảo cho quyết định đầu tư, vì ta phải xem xét nguồn gốc và các khoản mục chi tiết của nguồn lợi nhuận đó đến từ việc hoạt động kinh doanh có hiệu quả hay từ việc thanh lý và nhượng bán tài sản của công ty. Chính vì thế ta phải xem xét một cách kỹ lưỡng các chỉ số tài chính, để tránh đầu tư vào các công ty có giá trị thấp nhưng định giá cổ phiếu quá cao dẫn đến đầu tư lỗ vốn.

Với sự ra đời của vô số doanh nghiệp trên thương trường như ngày nay, vai trò của báo cáo tài chính ngày một quan trọng hơn và luôn được quan tâm nhiều nhất. Bởi nó thể hiện được tình hình tài chính, kinh doanh và các luồng tiền của doanh nghiệp. Nó là phương tiện trình bày khả năng sinh lời và thực trạng tài chính doanh nghiệp tới những người quan tâm.

Dựa trên nhu cầu đó, nhóm nghiên cứu đã thực hiện bài nghiên cứu về các chỉ số tài chính của hơn 200 công ty ở Hoa Kỳ để phân tích và làm rõ những vấn đề trên.

2. Giới thiệu đề tài

Cùng với tiến trình phát triển của hệ thống kế toán, hệ thống báo cáo tài chính cũng không ngừng được đổi mới để đáp ứng cho nhu cầu của người sử dụng thông tin trên báo cáo tài chính. Trong bối cảnh đó, các doanh nghiệp muốn đứng vững trên thương trường buộc phải nhanh chóng nắm bắt nhu cầu của người dùng và đổi mới để phù hợp với tiêu chí mà người dùng báo cáo tài chính đang hướng đến.

Phân tích báo cáo tài chính là công việc cần thiết và quan trọng đối với các nhà quản lý doanh nghiệp và nhà đầu tư. Các con số trên báo cáo tài chính giúp họ đưa ra nhiều quyết định quan trọng tác động đến lợi ích cá nhân cũng như lợi ích doanh nghiệp. Chẳng hạn như, đối với các nhà quản lý doanh nghiệp thì sẽ đề xuất ra nhiều giải pháp để kinh doanh có hiệu quả hơn, nâng cao năng lực và sức cạnh tranh của doanh nghiệp mình; đối với các nhà đầu tư, họ có thể dễ dàng hơn trong việc đưa ra quyết định có đầu tư hay không, nếu có đầu tư thì nên đầu tư bao nhiêu là hiệu quả nhất; hoặc các ngân hàng cũng

có thể đưa ra quyết định rằng có nên cho doanh nghiệp đó vay hay không thông qua việc phân tích báo cáo tài chính.

Thấy được tầm quan trọng và lợi ích của việc đọc và phân tích báo cáo tài chính, nhóm chúng em đã đưa ra quyết định rằng làm rõ hơn những con số trong báo cáo tài chính của doanh nghiệp thông qua việc “*Nghiên cứu về các chỉ số tài chính để đưa ra quyết định đầu tư*” bằng cách thu thập và phân tích dữ liệu báo cáo tài chính của hơn 200 công ty hàng đầu ở Hoa Kỳ để giúp các nhà đầu tư có cái nhìn chi tiết, rõ ràng hơn và dễ dàng đưa ra quyết định hơn về vấn đề đầu tư vào doanh nghiệp.

3. Mục tiêu nghiên cứu

3.1. Mục tiêu tổng quát

Việc nghiên cứu, phân tích báo cáo tài chính giúp những người quan tâm đến doanh nghiệp có một cái nhìn đúng đắn hơn về doanh nghiệp đó, đánh giá một cách thực tế nhất về doanh nghiệp, từ đó đưa ra các quyết định có lợi nhất cho bản thân. Những giải pháp vận hành được doanh nghiệp đưa ra sẽ hiệu quả hơn, nâng cao năng lực và sức cạnh tranh với các đối thủ; đối với các nhà đầu tư, giúp họ dễ dàng quyết định hơn về vấn đề đầu tư. Thông qua việc phân tích báo cáo tài chính các nhà quản lý doanh nghiệp cũng sẽ đưa ra được những chiến lược về huy động nguồn vốn và thu hút vốn vay từ phía các ngân hàng và nhà đầu tư một cách hiệu quả nhất.

3.2. Mục tiêu cụ thể

Bài toán 1: Dự đoán giá cổ phiếu của 200 công ty ở Hoa Kỳ.

Bài toán 2: Thông qua việc nghiên cứu báo cáo tài chính, phân tích khả năng nhận được nguồn vốn đầu tư từ các nhà đầu tư.

Bài toán 3: Thông qua việc nghiên cứu báo cáo tài chính, đánh giá kết quả kinh doanh của hơn 200 công ty ở Hoa Kỳ.

4. Đối tượng nghiên cứu

Sử dụng bộ dữ liệu bao gồm các thông tin công khai về danh sách báo cáo tài chính của hơn 200 công ty hàng đầu ở Hoa Kỳ do kaggle.com cung cấp.

5. Phương pháp nghiên cứu

5.1. Phương pháp phân lớp

Phân lớp dữ liệu là “*quá trình phân một dữ liệu vào một hay nhiều lớp (loại) đã cho trước nhờ một mô hình phân lớp. Mô hình này đã được xây dựng dựa trên một tập dữ liệu đã được gán nhãn trước đó. Quá trình gán nhãn cho một đối tượng dữ liệu chính là quá trình phân lớp*”.

Quá trình phân lớp dữ liệu gồm 2 bước chính:

Bước 1: Xây dựng mô hình phân lớp

- “Dữ liệu đầu vào: là dữ liệu mẫu đã được gán nhãn và tiền xử lý”

- “Các thuật toán phân lớp: cây quyết định, hàm số toán học,...”
- “Kết quả của bước này là mô hình phân lớp đã được huấn luyện (trình phân lớp)”

Bước 2.1: Đánh giá mô hình

- “Dữ liệu đầu vào: là một tập dữ liệu mẫu khác đã được gán nhãn và tiền xử lý. Tuy nhiên lúc đưa vào mô hình phân lớp, ta “lờ” đi thuộc tính đã được gán nhãn.”
- “Tính đúng đắn của mô hình sẽ được xác định bằng cách so sánh thuộc tính dán nhãn của dữ liệu đầu vào và kết quả phân lớp của mô hình.”

Bước 2.2: Phân lớp dữ liệu mới

- “Dữ liệu đầu vào: là dữ liệu “khuyết” thuộc tính cần dự đoán lớp “nhãn””
- “Mô hình sẽ tự động phân lớp (gán nhãn) cho các đối tượng dữ liệu này dựa vào những gì được huấn luyện ở bước 1.”

Phân loại bài toán phân lớp: nhiệm vụ của bài toán phân lớp là phân các đối tượng dữ liệu vào n lớp cho trước. Nếu:

- $n = 2$: thuộc bài toán phân lớp nhị phân
- $n > 2$: thuộc bài toán phân lớp đa lớp

Các phương pháp phân lớp được sử dụng trong bài

- *Phương pháp SVM (Support Vector Machine):* “là một thuật toán có giám sát, SVM nhận dữ liệu vào xem chúng như Vector trong không gian và phân loại chúng vào các lớp khác nhau bằng cách xây dựng một siêu phẳng trong không gian nhiều chiều là mặt phân cách các lớp dữ liệu. Để tối ưu kết quả phân lớp thì phải xác định siêu phẳng (Hyperlane) có khoảng cách đến các điểm dữ liệu (Margin) của tất cả các lớp xa nhất có thể. SVM có nhiều biến thể để phù hợp với nhiều bài toán phân loại khác nhau.”
- *Phương pháp Logistic Regression:* Logistic Regression “là phương pháp nhằm kiểm tra tính hiệu quả của mô hình phân lớp dữ liệu có đặc thù cụ thể, từ đó quyết định có sử dụng mô hình đó hay không. Một mô hình lý tưởng là một mô hình không quá đơn giản, không quá phức tạp, và không quá nhạy cảm với nhiễu”
- *Phương pháp cây quyết định (Decision Tree):* Cây quyết định là một sơ đồ hoặc biểu đồ giúp xác định một quá trình hành động hoặc hiển thị một xác suất thống kê. Biểu đồ được gọi là cây quyết định do nó giống với cây cùng tên, thường được phác thảo dưới dạng một biểu đồ thẳng đứng hoặc nằm ngang phân nhánh. Bắt đầu từ chính quyết định (được gọi là “nút”), mỗi “nhánh” của cây quyết định đại diện cho một quyết định, kết quả hoặc phản ứng có thể xảy ra. Các nhánh xa nhất trên cây đại diện cho kết quả cuối cùng của một con đường quyết định nhất định và được gọi là “lá”

5.2. Phương pháp phân cụm (Clustering)

Phân cụm dữ liệu là quá trình gom cụm/nhóm các đối tượng/dữ liệu có đặc điểm tương đồng vào các cụm/nhóm tương ứng. Trong đó:

- “Các đối tượng trong cùng một cụm sẽ có những tính chất tương tự nhau.”
- “Các đối tượng thuộc cụm/nhóm khác nhau sẽ có tính chất khác nhau.”

Đặc điểm:

- “Nhiệm vụ chính là tìm ra và đo đạc sự khác biệt giữa các đối tượng dữ liệu.”
- “Phân cụm thuộc nhóm phương pháp học không giám sát vì không biết trước được số nhóm.”
- “Một phương pháp phân cụm tốt là phương pháp tạo ra các cụm có chất lượng cao.”

6. Mô tả dữ liệu

Thuộc tính	Ý nghĩa	Mô tả
Ebitda Margins	Hệ số biên lợi nhuận trước thuế và lãi vay	Số thập phân
Profit Margins	Biên lợi nhuận	Số thập phân
Gross Margins	Biên lợi nhuận gộp	Số thập phân
Operating Cash flow	Dòng tiền từ hoạt động kinh doanh	Số tự nhiên
Revenue Growth	Tăng trưởng doanh thu	Số thập phân
Operating Margins	Mức lợi nhuận tạo ra từ hoạt động kinh doanh	Số thập phân
Ebitda	Thu nhập trước thuế, lãi vay, khấu hao	Số tự nhiên
Gross Profit	Lợi nhuận gộp	Số tự nhiên

Free Cash Flow	Dòng tiền tự do	Số tự nhiên
Current Price	Giá hiện hành	Số thập phân
Earnings Growth	Tăng trưởng lợi nhuận	Số thập phân
Current Ratio	Tỷ số thanh toán hiện hành	Số thập phân
Return On Assets	Tỷ số lợi nhuận trên tài sản	Số thập phân
Debt To Equity	Hệ số nợ trên vốn chủ sở hữu	Số thập phân
Return On Equity	Tỷ suất lợi tức trên vốn chủ sở hữu	Số thập phân
Total Cash	Tổng số tiền	Số tự nhiên
Total Debt	Tổng số nợ	Số tự nhiên
Total Revenue	Tổng doanh thu	Số tự nhiên
Total Cash Per Share	Tổng dòng tiền trên mỗi cổ phiếu	Số thập phân
Financial Currency	Đơn vị tiền tệ tài chính	Kí tự chữ
Revenue Per Share	Thu nhập trên mỗi cổ phiếu	Số thập phân
Quick ratio	Tỷ số thanh toán nhanh	Số thập phân
Quote Type	Loại báo giá	Kí tự chữ
Symbol	Kí hiệu	Kí tự chữ

Enterprise to revenue	Doanh thu	Số thập phân
Enterprise Ebitda	Lợi nhuận của doanh nghiệp	Số thập phân
Forward Eps	Lợi nhuận trên vốn cổ phần ước tính	Số thập phân
Shares Outstanding	Số phiếu đang lưu hành	Số tự nhiên
Book Value	Giá trị sổ sách	Số thập phân
Trailing Eps	Lợi nhuận trên vốn cổ phần hiện tại	Số thập phân
Price To Book	Chỉ số tài chính	Số thập phân
Held Percent Insiders	Tỷ lệ sở hữu cổ phần của cổ đông nội bộ	Số thập phân
Enterprise Value	Giá trị doanh nghiệp	Số tự nhiên
Earnings Quarterly Growth	Tăng trưởng thu nhập theo quý	Số thập phân
Peg Ratio	Tỷ số định giá cổ phiếu trên mức độ tăng trưởng	Số thập phân
Forward PE	Chỉ số ước tính của giá và lợi nhuận thu được trên mỗi cổ phiếu	Số thập phân
Market Cap	Vốn hoá thị trường	Số tự nhiên

	Định giá cổ phiếu	Kí tự chữ
	Sức khỏe tài chính	Kí tự chữ

Bảng 1. Mô tả dữ liệu

- Công thức sử dụng trong cột định giá cổ phiếu để xuất ra 2 giá trị là cao và thấp: =IF(định giá cổ phiếu dự trên tốc độ tăng trưởng>1;"Cao";"Thấp")
- Công thức sử dụng trong cột sức khỏe tài chính để xuất ra 2 giá trị tốt và không tốt: =IF(AND(3>tỷ số thanh toán hiện hành>1;tỷ suất sinh lời trên tài sản>0;tỷ số thanh toán nhanh>1;"tốt";"không tốt"))

Cơ sở lý luận:

- Định giá cổ phiếu:

Tỉ lệ PEG so sánh giữa tỉ lệ P/E với tốc độ tăng trưởng EPS kì vọng của nó.

- TH1: “Tỉ lệ PEG = 1, tăng trưởng thu nhập EPS của cổ phiếu đã được thị trường định giá đầy đủ vào trong giá của cổ phiếu.”
- TH2: “Tỉ lệ PEG > 1, cổ phiếu có thể đang bị định giá quá cao. Nói cách khác, mức tăng trưởng thu nhập mà thị trường kì vọng vào cổ phiếu đó cao hơn mức tăng trưởng thật sự mà cổ phiếu đó có thể tạo được.”
- TH3: “Tỉ lệ PEG < 1, cổ phiếu có thể nào đó đang bị định giá thấp hoặc thị trường đã không kì vọng công ty có thể đạt được tăng trưởng thu nhập giống như những dự báo mà công ty đưa ra.”

- Sức khỏe tài chính:

- “Một công ty có hệ số thanh toán hiện hành lớn hơn 1 sẽ có khả năng thanh toán hết các khoản nợ ngắn hạn do công ty không có lo ngại về khả năng thanh khoản trong ngắn hạn. Hệ số thanh toán hiện hành quá cao, trên 3, có thể cho thấy rằng công ty có thể trả các khoản nợ hiện có của mình gấp ba lần. Nó cũng

có thể là một dấu hiệu cho thấy công ty không quản lý hiệu quả các quỹ của mình.”

- “Nếu tỷ suất sinh lời trên tài sản lớn hơn 0, thì có nghĩa doanh nghiệp làm ăn có lãi. Tỷ số càng cao cho thấy doanh nghiệp làm ăn càng hiệu quả. Còn nếu tỷ số nhỏ hơn 0, thì doanh nghiệp làm ăn thua lỗ.”
- “Khả năng thanh toán nhanh > 1 . Khi hệ số lớn hơn hoặc bằng 1 thể hiện cho khả năng thanh toán ngay các khoản nợ ngắn hạn của doanh nghiệp nằm ở mức cao. Trong tình trạng này, đa số doanh nghiệp không gặp phải vấn đề trong việc thanh toán luôn các khoản nợ ngắn hạn.”
- “Khả năng thanh toán nhanh < 1 . Ngược lại, khi hệ số thanh toán nhanh nhỏ hơn 1 đồng nghĩa với khả năng thanh toán toàn bộ khoản nợ ngắn hạn trong thời gian ngắn của doanh nghiệp là không thể. Hay nói chính xác hơn, doanh nghiệp sẽ gặp vấn đề trong việc thanh toán nhanh chóng các khoản nợ ngắn hạn.”

CHƯƠNG II: QUY TRÌNH THỰC HIỆN VÀ KẾT QUẢ

1. Tiền xử lý

1.1. Tổng quan:

Tiền xử lý “là quá trình xử lý dữ liệu thô/gốc (*raw/original data*) nhằm cải thiện chất lượng dữ liệu (*quality of the data*) và do đó, cải thiện chất lượng của kết quả khai phá”

1.2. Quy trình thực hiện

Đầu tiên, Nhập file *financialdata* vào mục **File** sau đó tiến hành loại bỏ các cột giá trị không cần thiết (*totalcash, totalDebt, RevenuePerShare, Symbol, EnterpriseToRevenue, EnterpriseToEbitda, PriceToBook, HeldPercentInsiders, EnterpriseValue, EarningsQuarterlyGrowth*) thông qua **Select Columns**.

Quan sát dữ liệu trên bảng biểu bằng cách nói **Select Columns** vào **Data Table** thấy 226 instances, 26 feature (1.7% missing values)

Info

226 instances
26 features (1.7% missing values)
No target variable.
3 meta attributes (no missing values)

Variables

☒ Show variable labels (if present)
☒ Visualize numeric values
☒ Color by instance classes

Selection

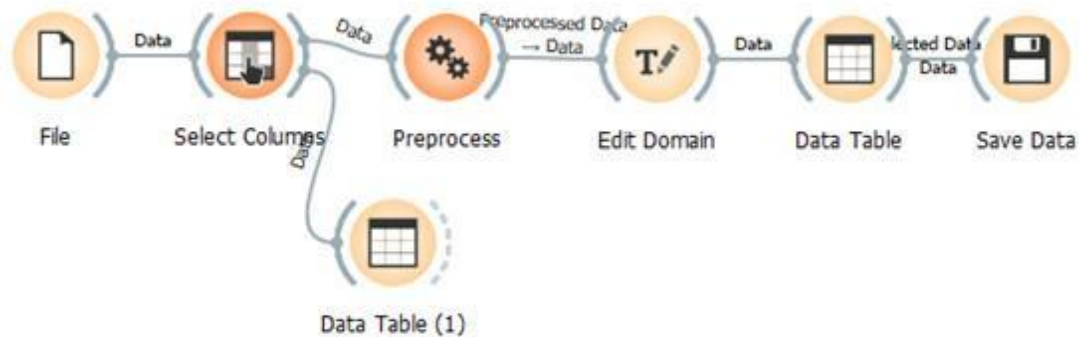
☐ Select full rows

Hình 1. Thông tin dữ liệu đầu vào

Xử lý các giá trị bị thiếu ở một số ô bằng cách lấy giá trị trung bình bằng chức năng **Preprocess**

Sau đó, Tiến hành đổi tên biến thành tiếng việt bằng **Edit domain**.
Xuất dữ liệu vừa xử lý ra **Data Table**

Lưu dữ liệu đã xử lý và đổi tên file thành BCTC các công ty ở Hoa Kỳ.



Hình 2. Mô hình tiền xử lý dữ liệu bằng Orange

2. Xử lý các bài toán

2.1. Bài toán Phân lớp dữ liệu

2.1.1. Các phương pháp dự đoán và quy trình cụ thể

** Các bước của quy trình SVM:*

Bước 1: Nhập dữ liệu huấn luyện vào orange.

Bước 2: Nối widget dữ liệu huấn luyện và SVM với Test and Score, sau đó nối widget vào Confusion Matrix để thực hiện đánh giá kết quả và đánh giá ma trận nhầm lẫn.

Bước 3: Nối dữ liệu huấn luyện vào SVM. Đồng thời nhập dữ liệu dùng để dự báo vào orange.

Bước 4: Liên kết SVM và dữ liệu dự báo với Predictions để đánh giá và phân loại dữ liệu đầu vào.

Bước 5: Xuất kết quả dự báo bằng Data Table.

** Các bước của quy trình Logistic Regression:*

Bước 1: Nhập dữ liệu huấn luyện vào orange.

Bước 2: Nối widget dữ liệu huấn luyện và Logistic Regression với Test and Score, sau đó nối widget vào Confusion Matrix để thực hiện đánh giá kết quả và đánh giá ma trận nhầm lẫn.

Bước 3: Nối dữ liệu huấn luyện vào Logistic Regression. Đồng thời nhập dữ liệu dùng để dự báo vào orange.

Bước 4: Liên kết Logistic Regression và dữ liệu dự báo với Predictions để đánh giá và phân loại dữ liệu đầu vào.

Bước 5: Xuất kết quả dự báo bằng Data Table.

** Các bước của quy trình Decision Tree:*

Bước 1: Nhập dữ liệu huấn luyện vào orange.

Bước 2: Nối widget dữ liệu huấn luyện và Tree với Test and Score, sau đó nối widget vào Confusion Matrix để thực hiện đánh giá kết quả và đánh giá ma trận nhầm lẫn.

Bước 3: Nối dữ liệu huấn luyện vào Tree. Đồng thời nhập dữ liệu dùng để dự báo vào orange.

Bước 4: Liên kết Tree và dữ liệu dự báo với Predictions để đánh giá và phân loại dữ liệu đầu vào.

Bước 5: Xuất kết quả dự báo bằng Data Table.

2.1.2. Xử lý các bài toán phân lớp

a. Phân tách dữ liệu

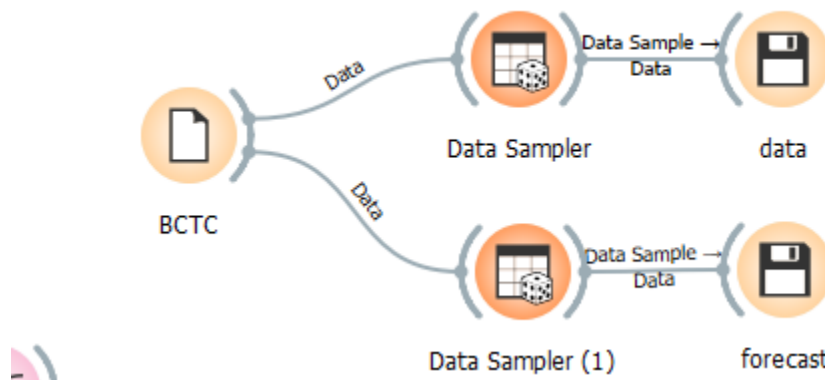
Trong bài này, nhóm nghiên cứu đã tách dữ liệu khảo sát ban đầu thành 2 file riêng biệt để thực hiện việc phân lớp dữ liệu như sau: Sử dụng 70% dữ liệu ban đầu để làm dữ liệu mẫu huấn luyện mô hình phân lớp dữ liệu. Và sử dụng 30% dữ liệu còn lại để làm dữ liệu dự báo cho nghiên cứu.

Quy trình như sau:

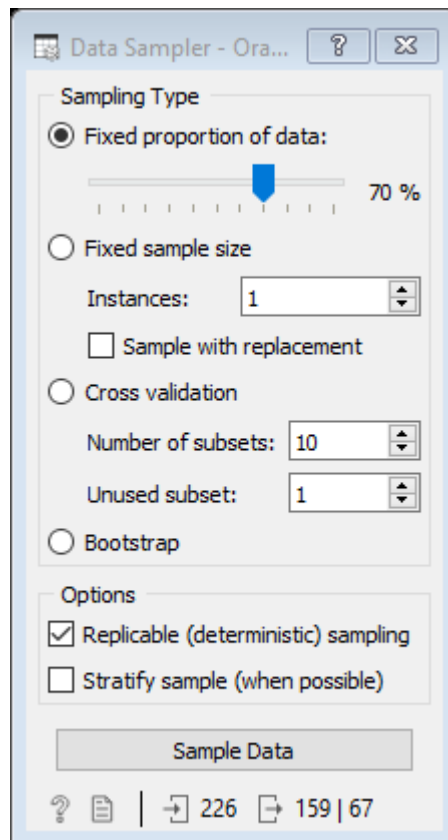
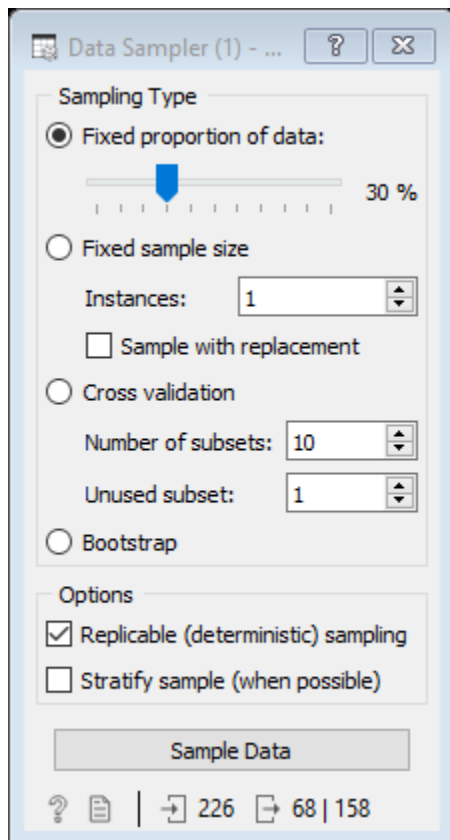
Bước 1: Chọn flie BCTC các công ty Hoa Kỳ

Bước 2: xử dụng data sampler để trích xuất dữ liệu thành 70% và 30%

Bước 3: chọn save data để Lưu dữ liệu trích xuất 70% với tên data
Tiếp tục chọn save data để lưu dữ liệu 30% với tên forecast.



Hình 3 : Mô hình phân tách dữ liệu của bài toán 1



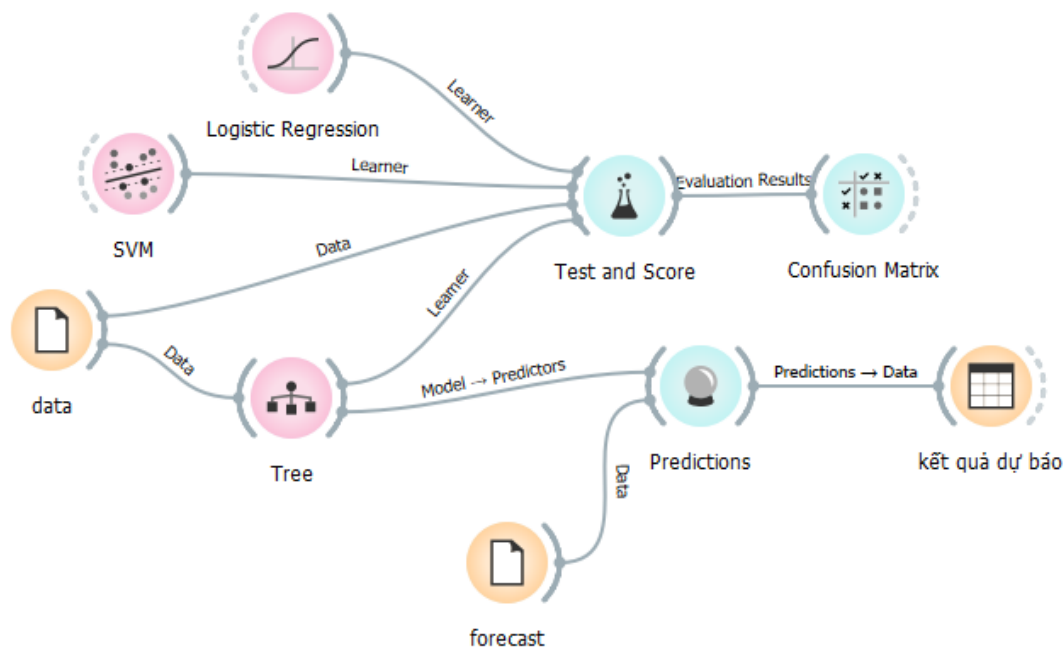
b. Xử lý bài toán 1: Dự đoán giá cổ phiếu

- Xây dựng mô hình:

Bước 1: Chọn dữ liệu File Data.xlsx đã trích xuất 70% dữ liệu ban đầu đã trích xuất ở trên. Trong đó biến Target là Định giá cổ phiếu, skip biến Sức khỏe tài chính vì không liên quan đến kết quả bài toán 1 và để tránh xuất hiện lỗi.

Bước 2: Dùng 3 phương pháp: SVM, Tree, Logistic Regression tiến hành phân tích dự đoán cổ phiếu của các công ty ở Mỹ thuộc khoảng nào & đánh giá hiệu quả các phương pháp.

Bước 3: Chọn phương pháp được đánh giá tốt nhất, dùng phương pháp đó dự báo cho dữ liệu File “Forecast” . (Chọn dữ liệu từ file Forecast đã trích xuất 30% từ dữ liệu ban đầu, biến Target biến Định giá cổ phiếu, skip biến Sức khỏe tài chính) sau đó thì cho ra bảng kết quả dự báo.

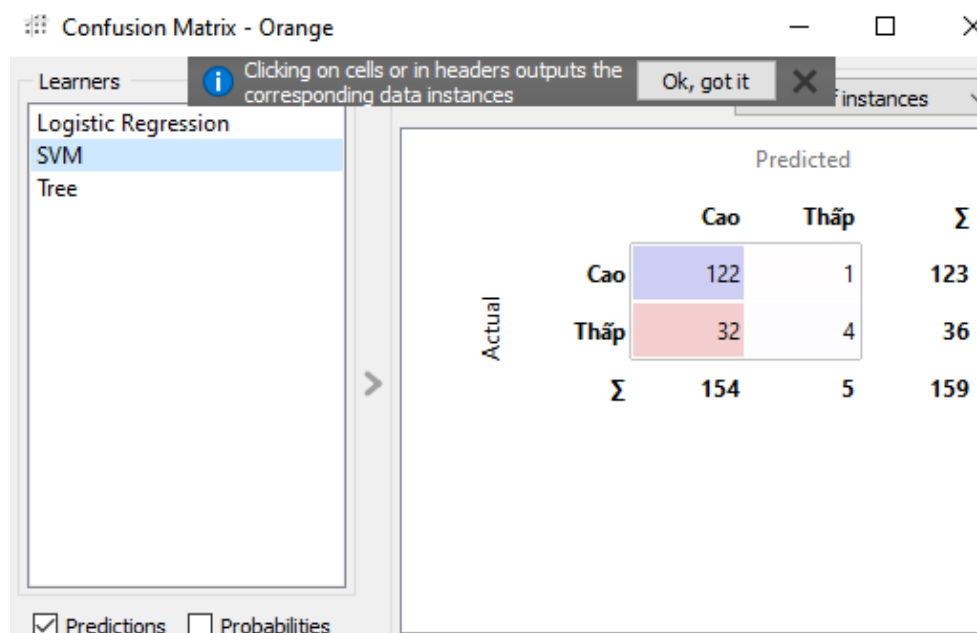
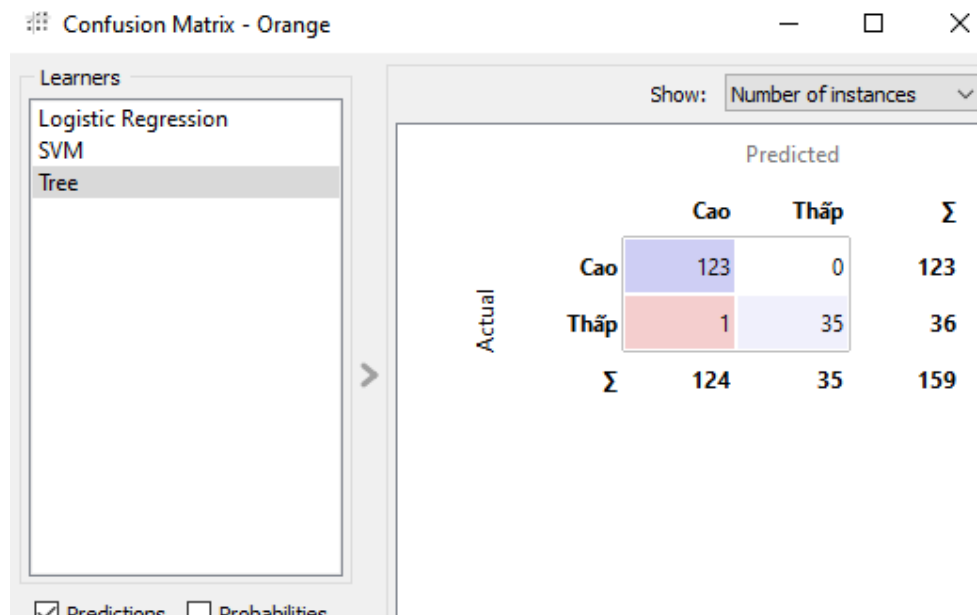


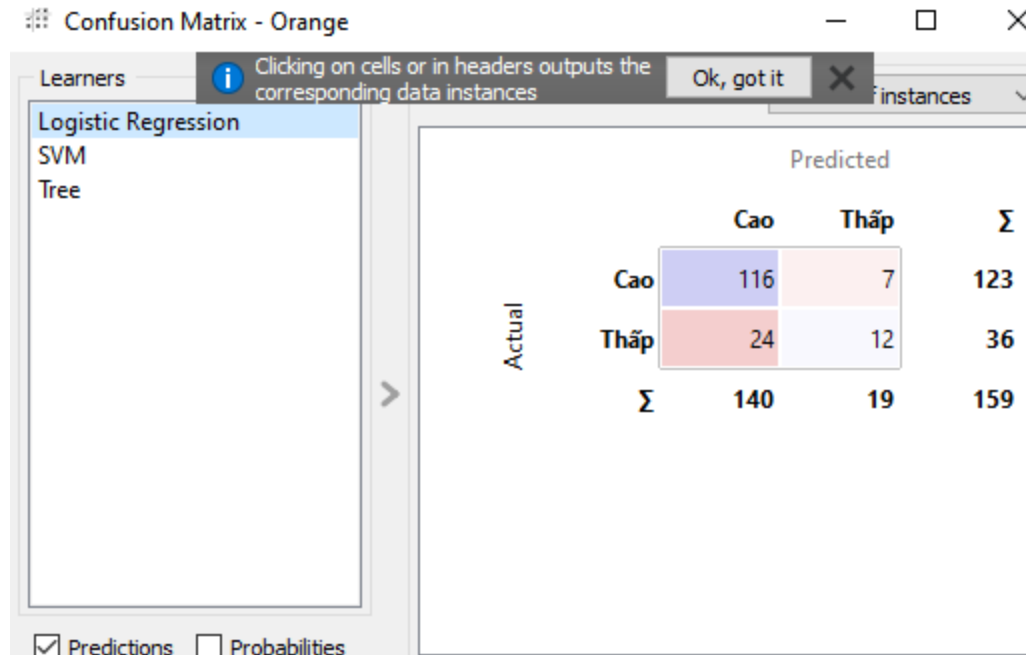
Hình 4 : Mô hình bài toán 1

**Kết quả và đánh giá:*

+ **Theo ma trận nhầm lẫn**

- Sai lầm loại 2 : dự báo Thấp thực tế là Cao dẫn đến đánh giá sai nghiêm trọng khả năng về dự đoán giá cổ phiếu
- Ta thấy Tree = 1 < Logistic Regression (=24) và SVM (= 32) nên ta chọn Tree làm phương pháp đánh giá cho ra mô hình hiệu quả nhất vì chỉ số sai càng thấp mô hình phân lớp sẽ càng tốt.





Bảng 2 : Kết quả Confusion Matrix

--- > Để chắc chắn hơn chúng ta sẽ sử dụng thêm Test & Score để kiểm tra

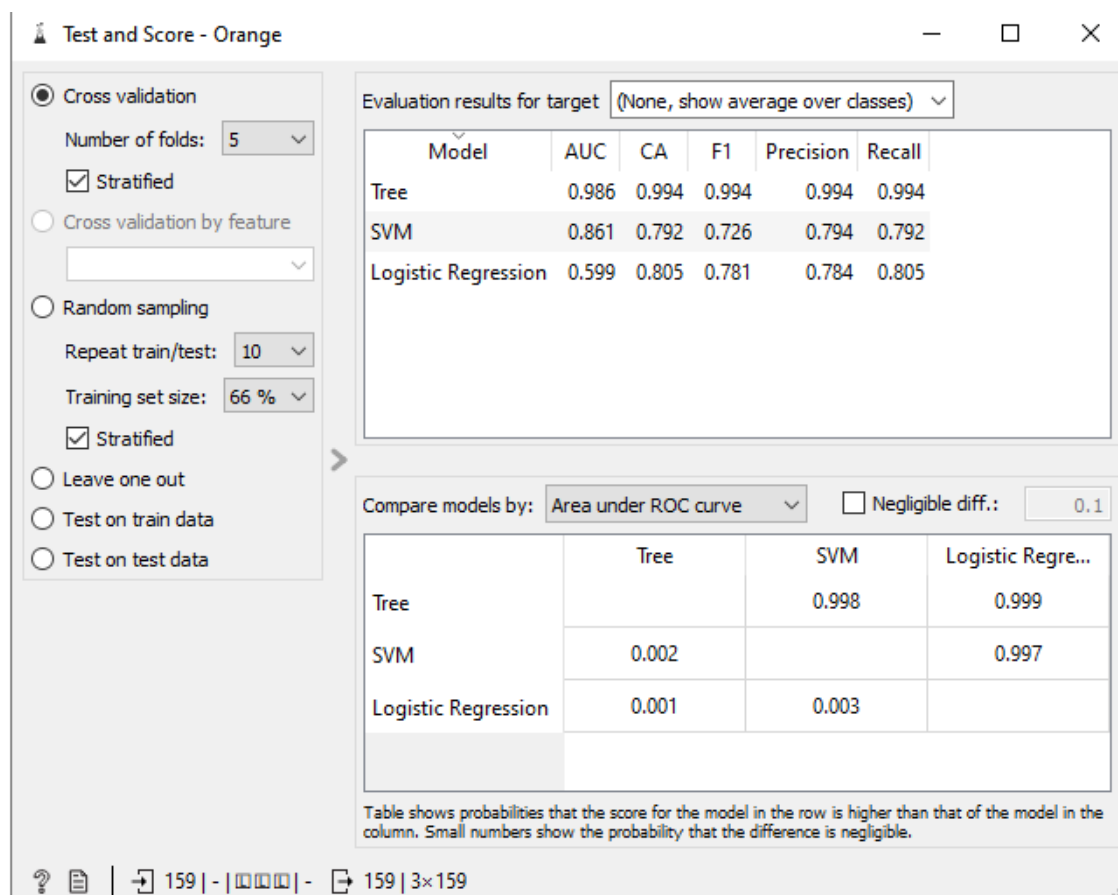
- **Test and Score**

Tại mục Number of folds ta chọn 5

Xét các chỉ số :

- “AUC (Area Under the Curve) là phần diện tích nằm dưới đường cong ROC .”
- “CA (Tính chính xác) là tỷ lệ số mẫu được phân lớp đúng trong toàn bộ tập dữ liệu.”
- “Precision (Độ chính xác) cho biết số mẫu m được phân vào lớp i thì có tỷ lệ bao nhiêu mẫu có đúng.”
- “Recall (Độ phủ) còn gọi là độ phủ hay độ nhạy.”
- “F1-score là giá trị trung bình điều hòa của hai độ đo Precision và Recall.”

Tất cả các chỉ số trên càng tiến về 1 thì mô hình càng tốt. Như vậy chúng ta dễ dàng nhận ra rằng phương pháp phân lớp **Tree** là phương pháp tốt nhất trong 3 phương pháp trên. Vì chỉ số AUC của Tree = 0.986 Lớn hơn chỉ số AUC của SVM (=0.861) và Logistic Regression (= 0.599).



Bảng 3 : Kết quả Test & Score của bài toán 1

Như vậy bài toán 1 sẽ dùng phương pháp Tree để dự đoán giá cổ phiếu cho ra kết quả tốt nhất.

- Kết quả dự báo :

	định giá cổ phiếu	Tên công ty	Ngành	Mã CK
31	Thấp	General Motors...	Auto Manufact...	GM
32	Cao	Target Corporat...	Discount Stores	TGT
33	Thấp	Chevron Corpo...	Oil & Gas Integ...	CVX
34	Cao	Carrier Global C...	Building Produ...	CARR
35	Cao	Digital Realty Tr...	REIT—Office	DLR
36	Cao	Equinix, Inc.	REIT—Specialty	EQIX
37	Thấp	Baker Hughes ...	Oil & Gas Equip...	BKR
38	Cao	Public Storage	REIT—Industrial	PSA
39	Cao	Dominion Ener...	Utilities—Regul...	D
40	Cao	Ecolab Inc.	Specialty Chem...	ECL
41	Cao	Archer-Daniels-...	Farm Products	ADM
42	Cao	ResMed Inc.	Medical Instru...	RMD
43	Cao	Dollar General ...	Discount Stores	DG
44	Thấp	Halliburton Co...	Oil & Gas Equip...	HAL
45	Cao	DBA Sempra	Utilities—Divers...	SRE
46	Cao	Cadence Desig...	Software—Appl...	CDNS
47	Cao	ServiceNow, Inc.	Software—Appl...	NOW
48	Cao	S&P Global Inc.	Financial Data ...	SPGI
49	Cao	Eaton Corporati...	Specialty Indust...	ETN

Bảng 4 : Kết quả dự báo của bài toán 1

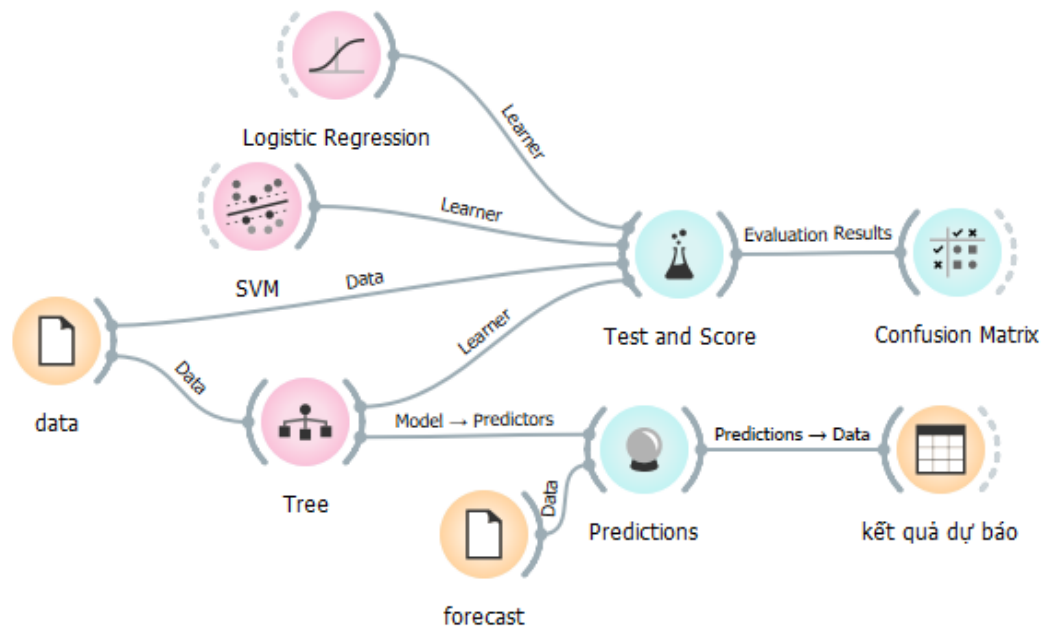
c. Xử lý bài toán 2 : Phân tích khả năng các công ty này nhận được thêm nguồn vốn đầu tư.

- *Xây dựng mô hình:*

Bước 1: Chọn dữ liệu File Data.xlsx đã trích xuất 70% dữ liệu ban đầu đã trích xuất ở bài toán 1. Trong đó biến Target là Sức khỏe tài chính và skip biến Định giá cổ phiếu vì không liên quan đến kết quả của bài toán 2 và tránh gây ra lỗi.

Bước 2: Dùng 3 phương pháp: SVM, Tree, Logistic Regression tiến hành phân tích khả năng cho vay vốn từ các ngân hàng ở Mỹ thuộc khoảng nào & đánh giá hiệu quả các phương pháp.

Bước 3: Chọn phương pháp được đánh giá tốt nhất, dùng phương pháp đó dự báo cho dữ liệu File “Forecast” . (Chọn dữ liệu từ file Forecast đã trích xuất 30% từ dữ liệu ban đầu, biến Target là Sức khỏe tài chính, skip biến Định giá cổ phiếu) sau đó thì cho ra bảng kết quả dự báo.



Hình 5 : Mô hình xây dựng bài toán 2

- Đánh giá và kết quả:

- Theo ma trận nhầm lẫn

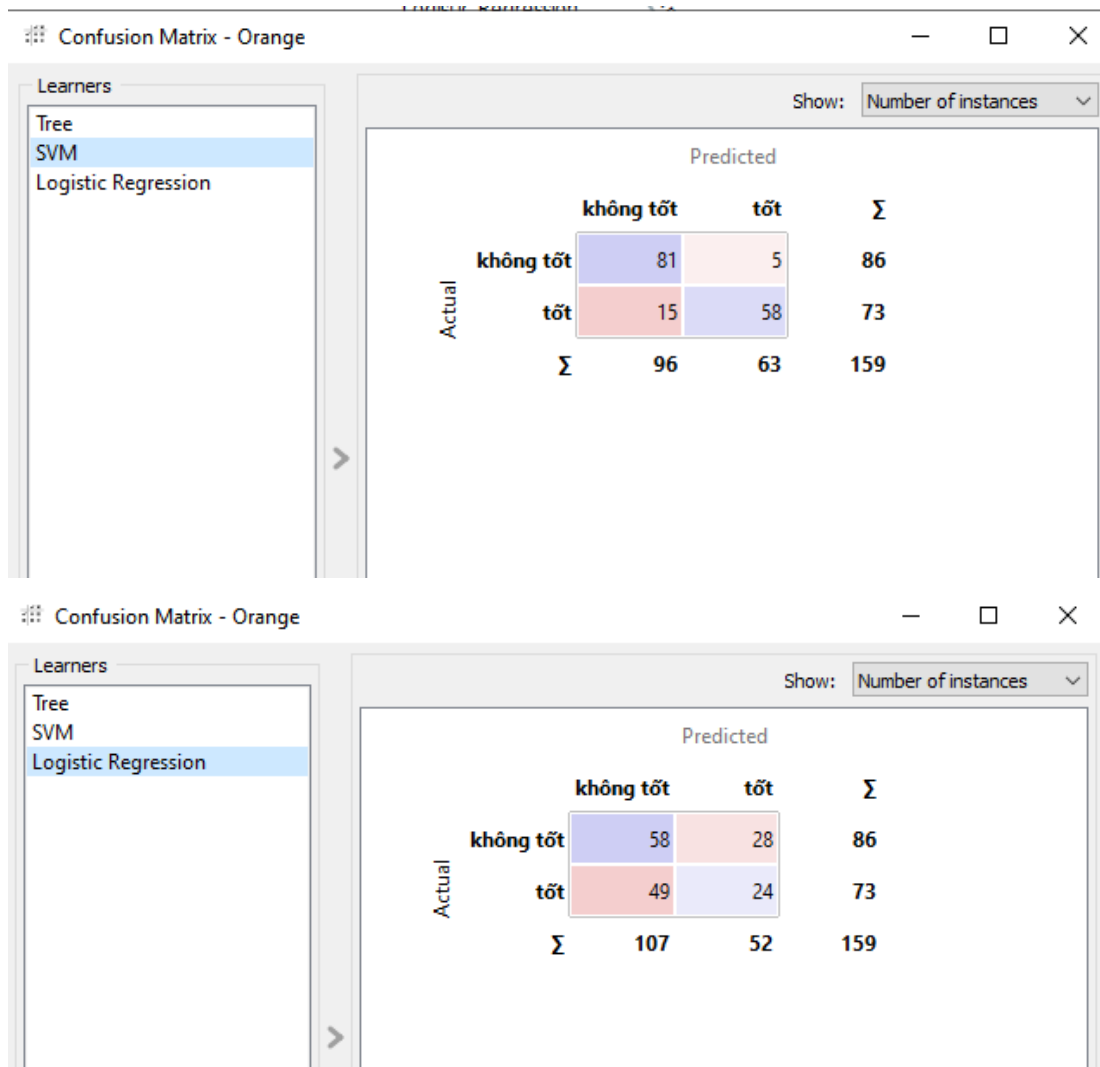
+ Sai lầm loại 2 : dự báo là Tốt nhưng thực tế là Không tốt dẫn đến đánh giá sai nghiêm trọng khả năng được nhận thêm nguồn vốn đầu tư của các công ty Hoa Kỳ.

+ Ta thấy Tree = 0 < Logistic Regression (=49) và SVM (= 15) nên ta chọn Tree làm phương pháp đánh giá cho ra mô hình hiệu quả nhất vì chỉ số sai càng thấp mô hình phân lớp sẽ càng tốt.

Confusion Matrix - Orange

Show: Number of instances

		Predicted		Σ
		không tốt	tốt	
Actual	không tốt	84	2	86
	tốt	0	73	73
Σ		84	75	159



Bảng 5: Kết quả Confusion Matrix của bài toán

Để chắc chắn hơn chúng ta sẽ sử dụng thêm Test & Score để kiểm tra

- **Test and Score**

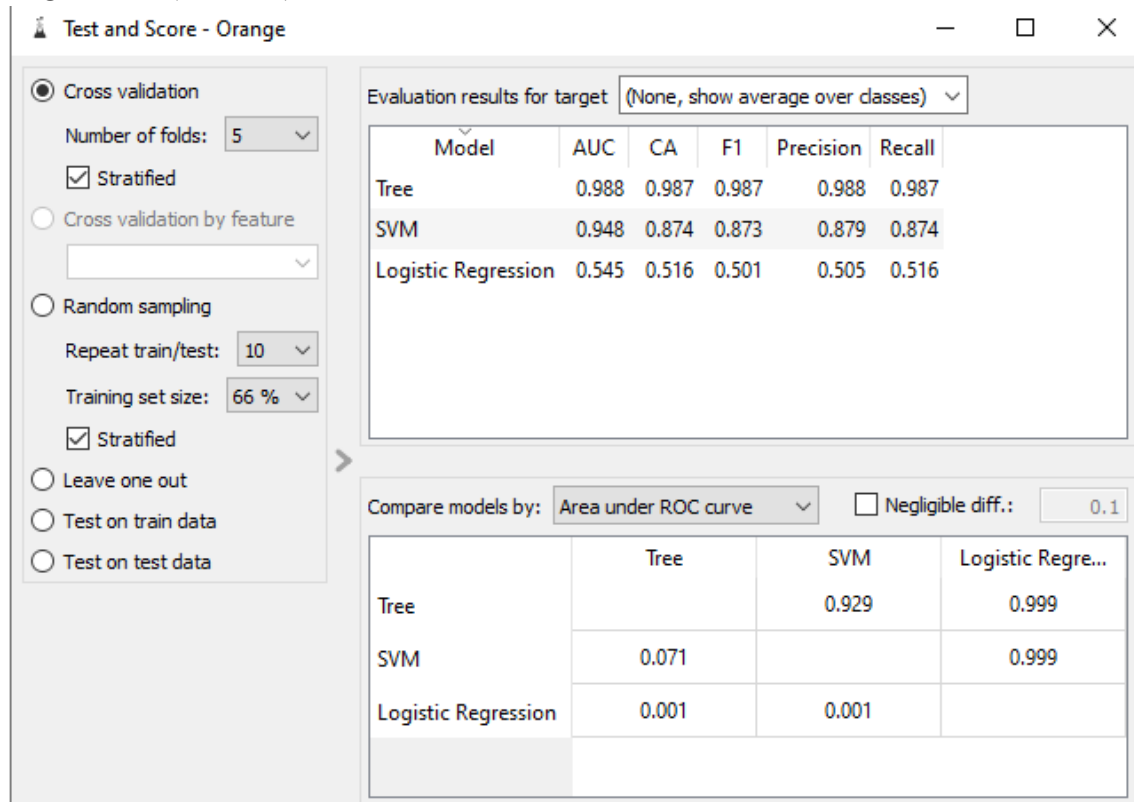
Tại mục Number of folds ta chọn 5

Xét các chỉ số :

- “AUC (Area Under the Curve) là phần diện tích nằm dưới đường cong ROC .”
- “CA (Tính chính xác) là tỷ lệ số mẫu được phân lớp đúng trong toàn bộ tập dữ liệu.”
- “Precision (Độ chính xác) cho biết số mẫu m được phân vào lớp i thì có tỷ lệ bao nhiêu mẫu có đúng.”
- “Recall (Độ phủ) còn gọi là độ phủ hay độ nhạy.”
- “F1-score là giá trị trung bình điều hòa của hai độ đo Precision và Recall.”

Tất cả các chỉ số trên càng tiến về 1 thì mô hình càng tốt. Như vậy chúng ta dễ dàng nhận ra rằng phương pháp phân lớp **Tree** là phương pháp tốt nhất trong 3 phương pháp trên vì

chỉ số AUC của Tree = 0.988 lớn hơn chỉ số AUC của SVM (= 0.948) và Logistic Regression (= 0.545)



Bảng 6 : Kết quả Test & Score của bài toán 2

Kết luận : Như vậy sẽ dùng phương pháp Tree để dự báo khả năng vay vốn của các công ty ở Mỹ cho bài toán 2.

- Kết quả dự báo như sau :

	sức khỏe tài chính	Tên công ty	Ngành	Mã CK
1	tốt	Visa Inc.	Credit Services	V
2	tốt	Nucor Corporat...	Steel	NUE
3	tốt	Vertex Pharmac...	Biotechnology	VRTX
4	không tốt	Parker-Hannifi...	Specialty Indust...	PH
5	tốt	Centene Corpo...	Healthcare Plans	CNC
6	không tốt	SBA Communi...	REIT—Specialty	SBAC
7	tốt	Monster Bevera...	Beverages—No...	MNST
8	tốt	Analog Devices...	Semiconductors	ADI
9	tốt	Welltower Inc.	REIT—Healthca...	WELL
10	tốt	Corteva, Inc.	Agricultural Inp...	CTVA
11	không tốt	Prologis, Inc.	REIT—Industrial	PLD
12	không tốt	Home Depot, I...	Home Improve...	HD
13	tốt	Baxter Internati...	Medical Instru...	BAX
14	tốt	Cisco Systems, ...	Communicatio...	CSCO
15	không tốt	Salesforce.com ...	Software—Appl...	CRM
16	không tốt	NextEra Energy,...	Utilities—Regul...	NEE
17	tốt	Deere & Comp...	Farm & Heavy ...	DE
18	không tốt	Constellation B...	Beverages—Wi...	STZ
19	tốt	Moderna, Inc.	Biotechnology	MRNA
20	không tốt	The Travelers C...	Insurance—Pro...	TRV
21	không tốt	Altria Group, Inc.	Tobacco	MO

Bảng 7 : Kết quả dự báo của bài toán 2

2.2. Bài toán phân cụm

- Xử lý bài toán 3: Đánh giá kết quả kinh doanh của 200 công ty ở Mỹ.

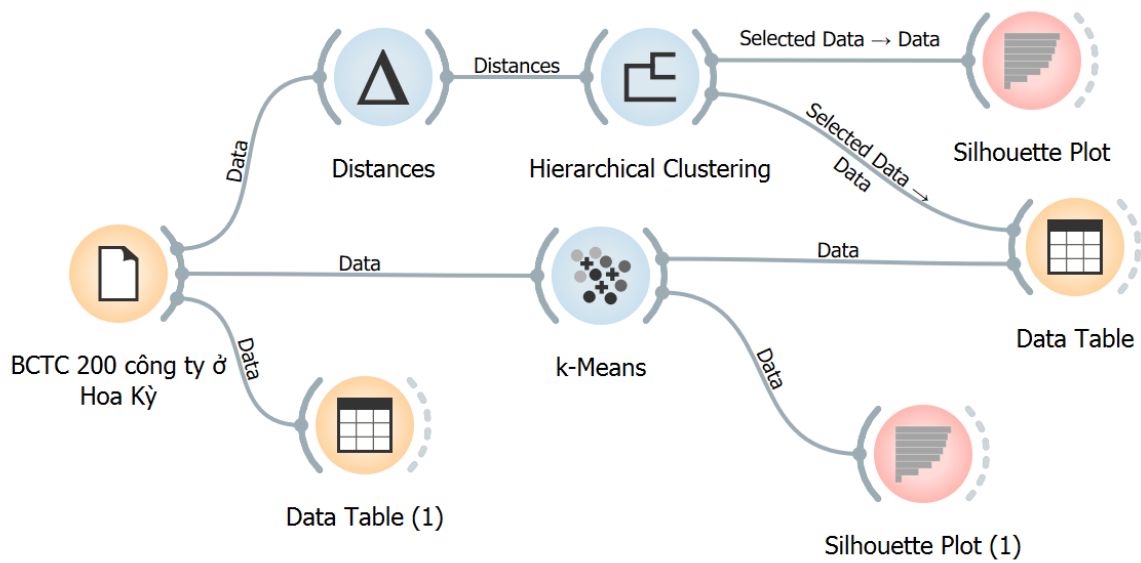
*Các phương pháp phân cụm:

Để thực hiện phân cụm bộ dữ liệu nhóm sử dụng hai phương pháp chính là Hierarchical clustering và K-means.

+ “Đối với phương pháp Hierarchical clustering nhóm tiến hành tính khoảng cách giữa các phần tử bằng Distance rồi quan sát dữ liệu được phân cụm với số cụm từ 2 đến 5 và quan sát trên Silhouette Plot.”

+ “Đối với phương pháp K-means nhóm quan sát chỉ số Silhouette trung bình khi phân dữ liệu từ 2 đến 5 cụm, chọn số cụm có chỉ số Silhouette tốt, phù hợp với số lượng biến có sẵn trên bộ dữ liệu và quan sát trên Silhouette Plot.”

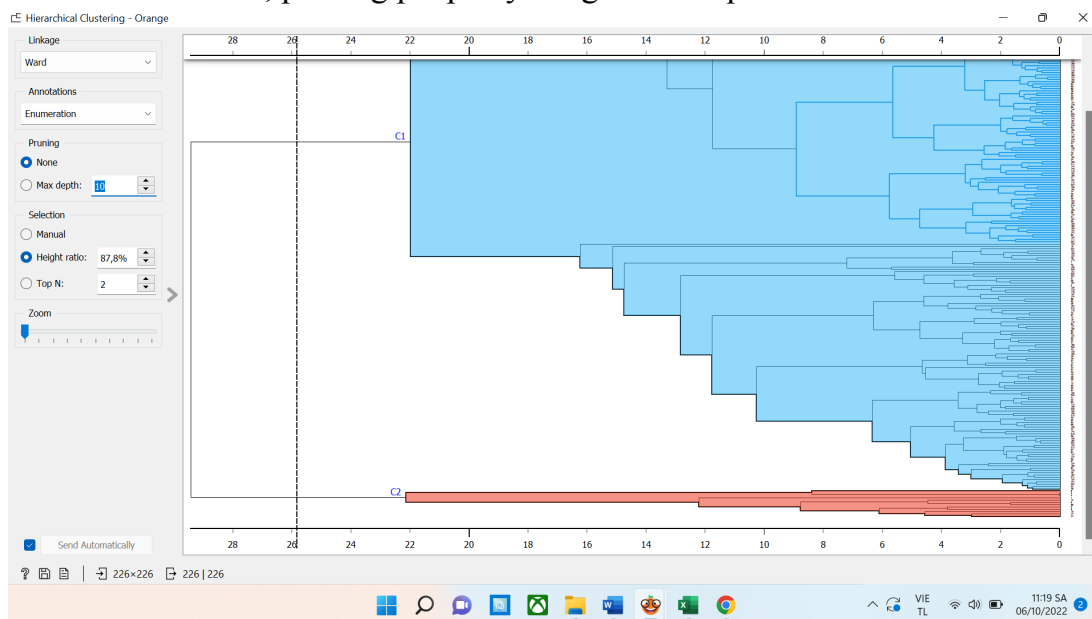
2.2.1. Qui trình cụ thể



Hình 6. Mô hình xây dựng bài toán 3

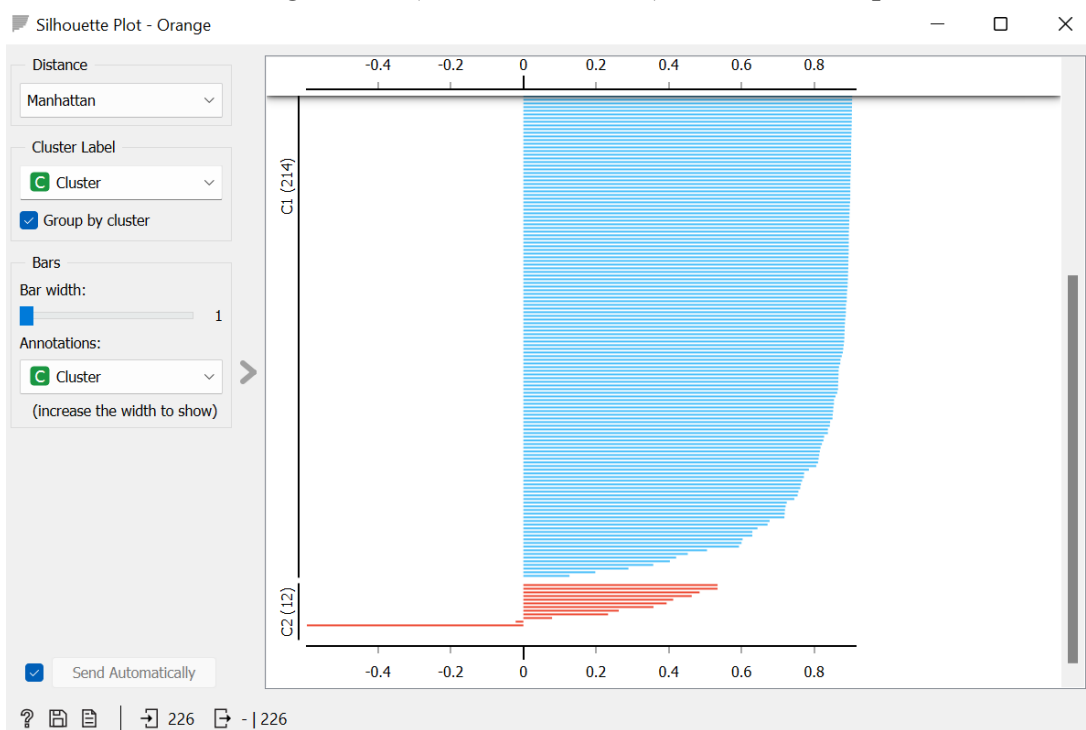
Về Phương pháp Hierarchical Clustering:

- Muốn thực hiện phương pháp này trước tiên cần có ma trận khoảng cách (Distances).
- Chọn **Hierarchical Clustering** để tiến hành phân cụm. Dựa vào biểu đồ và chỉ số Silhouette để tiến hành đánh giá. Sau khi xem xét và so sánh chỉ số Silhouette ở cả 4 phương pháp, ta nhận thấy rằng sử dụng **Ward** chia thành 2 cụm biểu đồ trông bắt mắt hơn, phương pháp này cũng cho kết quả tốt nhất.



Hình 7. Kết quả phân cụm theo phương pháp Hierarchical Clustering

- Để thấy rõ phương pháp **Ward** là tốt nhất, ta quan sát chỉ số Silhouette của cụm thứ nhất tương đối tốt (tỉ lệ trên 0.5 cao), cụm thứ 2 đa phần từ 0 đến 0.5, dưới 0 ít

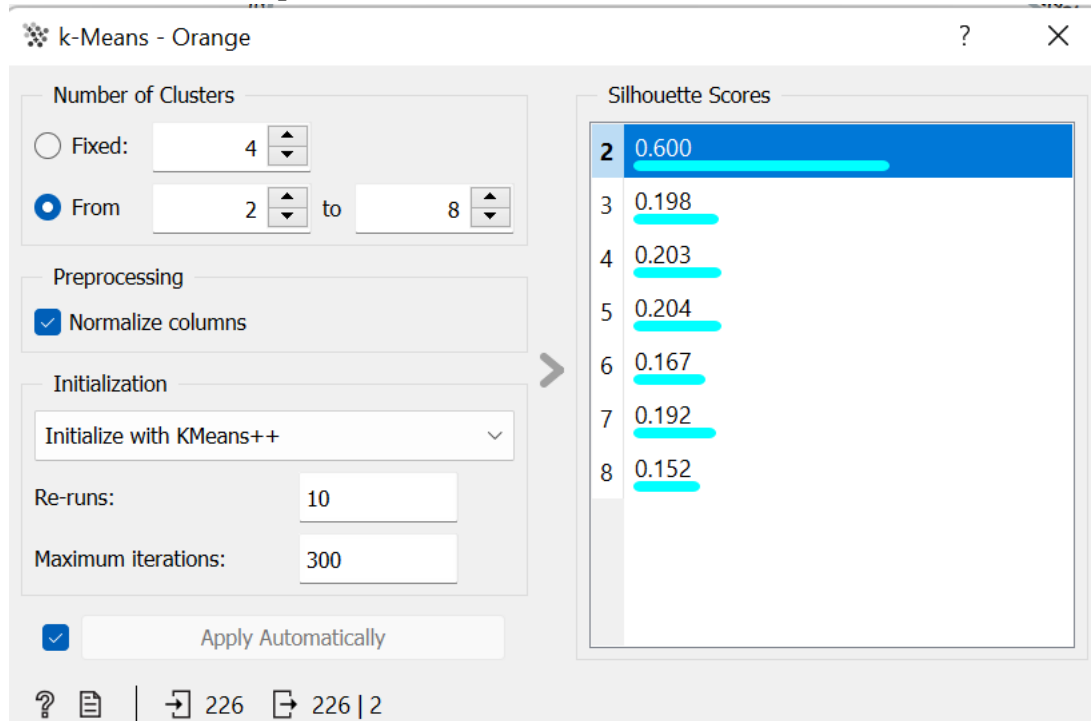


Hình 8. Chỉ số Silhouette của phương pháp Hierarchical Clustering

Về phương pháp K-Means:

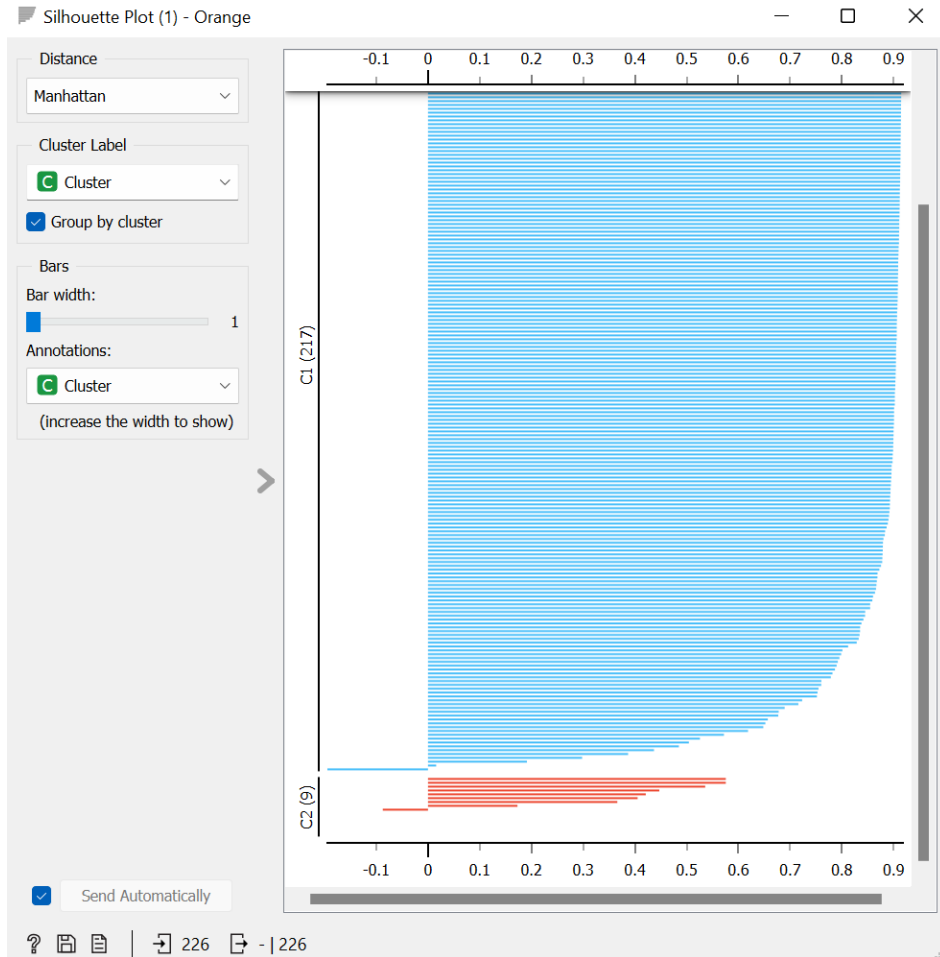
- Click chuột vào **K-Means**, xuất hiện hai lựa chọn là **fixed** và **from**, ta chọn **from** (vì chưa biết chính xác phân bao nhiêu cụm).
- Dựa vào bảng **Silhouette Scores**, về nguyên tắc điểm **Silhouette** này là một giá trị thuộc khoảng từ -1 đến 1, giá trị này càng lớn thì kết quả phân cụm càng tốt.

⇒ Ta nhận thấy phân thành 2 cụm có chỉ số Silhouette cao nhất là 0.600, đáng tin cậy nhất, do đó ta chọn phân thành 2 cụm là tốt nhất.



Hình 9. Kết quả phân cụm theo phương pháp K-Means

- Để thấy rõ lựa chọn phân thành 2 cụm là tốt nhất, ta quan sát chỉ số **Silhouette** cụm thứ nhất tương đối tốt (tỉ lệ trên 0.5 cao, tỉ lệ dưới 0 ít), cụm thứ 2 đa phần từ 0.1 đến 0.55, dưới 0 ít.



Hình 10. Chỉ số Silhouette của phương pháp K-Means

⇒ Dựa vào phương pháp K-Means và Hierarchical Clustering, ta kết luận phân thành 2 cụm là tốt nhất.

Kết quả sau khi phân cụm như sau:

Data Table - Orange

Info
226 instances (no missing data)
26 features
No target variable.
4 meta attributes

Variables
☒ Show variable labels (if present)
☐ Visualize numeric values
☒ Color by instance classes

Selection
☒ Select full rows

Restore Original Order

☒ Send Automatically

	Tên công ty	Ngành	Mã CK	Cluster	ợi nhuận trước thi	Biên lợi nhuận	Biên lợi
1	Apple Inc.	Consumer Elec...	AAPL	C2	0.33891	0.26579	
2	Microsoft Corp...	Software—Infr...	MSFT	C2	0.49123	0.38498	
3	Alphabet Inc.	Internet Conte...	GOOG	C2	0.35377	0.29512	
4	Alphabet Inc.	Internet Conte...	GOOGL	C2	0.35377	0.29512	
5	Amazon.com, I...	Internet Retail	AMZN	C2	0.12595	0.07101	
6	Tesla, Inc.	Auto Manufact...	TSLA	C1	0.15512	0.07403	
7	NVIDIA Corpor...	Semiconductors	NVDA	C1	0.40203	0.33806	
8	Meta Platforms...	Internet Conte...	FB	C2	0.46401	0.33384	
9	UnitedHealth G...	Healthcare Plans	UNH	C2	0.08662	0.05523	
10	Visa Inc.	Credit Services	V	C1	0.69612	0.51592	
11	Johnson & Joh...	Drug Manufact...	JNJ	C1	0.34803	0.22264	
12	Walmart Inc.	Discount Stores	WMT	C2	0.0639	0.02387	
13	Procter & Gam...	Household & P...	PG	C1	0.27022	0.1852	
14	Exxon Mobil C...	Oil & Gas Integ...	XOM	C2	0.16556	0.08259	
15	Mastercard Inc...	Credit Services	MA	C1	0.58218	0.46002	
16	Home Depot, I...	Home Improve...	HD	C1	0.17396	0.10791	
17	Chevron Corpo...	Oil & Gas Integ...	CVX	C1	0.21459	0.10041	
18	Pfizer, Inc.	Drug Manufact...	PFE	C2	0.39382	0.28684	
19	AbbVie Inc.	Drug Manufact...	ABBV	C1	0.50287	0.20538	
20	Eli Lilly and Co...	Drug Manufact...	LLY	C1	0.34789	0.21515	
21	Coca-Cola Co...	Beverages—No...	KO	C1	0.35199	0.23314	
22	Broadcom Inc.	Semiconductors	AVGO	C1	0.53639	0.24539	
23	Walt Disney Co...	Entertainment	DIS	C1	0.12761	0.02959	
24	Costco Wholes...	Discount Stores	COST	C1	0.04635	0.02543	
25	Pepsico, Inc.	Beverages—No...	PEP	C1	0.18109	0.10617	

Bảng 8: Bảng kết quả phân cụm của bài toán 3

Sau khi phân cụm, ta có thể dễ dàng nhận thấy được rằng cụm 1 chiếm phần đa số là định giá cổ phiếu **cao**- sức khỏe tài chính **không tốt**, cụm 2 chiếm phần đa số là định giá cổ phiếu **cao**- sức khỏe tài chính **tốt**.

CHƯƠNG III: KẾT LUẬN

1. Kết luận

Với nền kinh tế thị trường cũng như xu thế hội nhập quốc tế ngày càng phát triển mạnh mẽ đòi hỏi các doanh nghiệp trong quá trình hoạt động và phát triển phải tiến hành phân tích báo cáo tài chính. Báo cáo tài chính đóng một vai trò vô cùng quan trọng trong quá trình phân tích cơ bản một cách toàn diện tình hình tài chính của công ty.

Qua quá trình nghiên cứu và tìm hiểu về các chỉ số tài chính, phân tích tình hình tài chính của 200 công ty tại Hoa Kỳ, nhằm mục đích để các ban lãnh đạo, các nhà đầu tư và các bên cho vay có góc nhìn toàn diện về tình hình kinh tế tài chính, kinh doanh, về luồng tiền và khả năng chi trả của doanh nghiệp. Trước khi các nhà đầu tư và các bên cho vay quyết định đầu tư hay chỉ một khoản vay cho công ty, họ phải tìm hiểu kỹ về tình hình hoạt động của công ty, về các khoản nợ sẵn có hay các khoản lợi nhuận mà công ty kiếm được thông qua báo cáo tài chính để từ đó đưa ra các lựa chọn hình thức kinh doanh và các quyết định về kinh tế, đặc biệt là quyết định có nên đầu tư để phát triển hay cho công ty vay vốn hay không, đồng thời thông qua các chỉ số về “sức khỏe tài chính” để phân tích tình hình hoạt động ở các ngành từ đó so sánh và đưa ra kết luận về kết quả hoạt động của các công ty.

Ở bài nghiên cứu này, nhóm chúng em đã tập trung phân tích dữ liệu từ báo cáo tài chính và làm rõ những lý luận về phân tích tài chính. Nhóm đã tiến hành thu thập thông tin về báo cáo tình hình tài chính của hơn 200 công ty ở Hoa Kỳ trên nguồn kaggle.com. Thông qua những dữ liệu đó nhóm đã tiến hành nghiên cứu sâu hơn những dữ liệu về chỉ số tài chính và cho ra dự báo về giá cổ phiếu, dự báo về sức khỏe tài chính và đánh giá kết quả hoạt động ở hơn 200 công ty hàng đầu của Hoa Kỳ. Từ các kết quả phân tích trên không chỉ giúp cho ban lãnh đạo trong nội bộ của công ty nhận thấy được rõ tình hoạt động và khả năng tài chính hiện thực của công ty để đưa ra các quyết định về sự phát triển trong tương lai của công ty cũng như hướng đi thích hợp cho các dự định kinh doanh và dự án trong tương lai một cách thích hợp để thu hút các nhà đầu tư một cách hiệu quả nhất. Bên cạnh đó các số liệu phân tích được sẽ giúp cho những người bên ngoài công ty cần đến số liệu tài chính này để đưa ra các quyết định đầu tư có thể nhìn thấy được khả năng tài chính của công ty một cách rõ ràng, chi tiết, khách quan và độ tin cậy cao nhất để đưa ra các quyết định đầu tư quan trọng. Chỉ số tài chính còn giúp cho ta thấy được sự chênh lệch về thực lực tài chính giữa các công ty cùng ngành, sự so sánh đó giúp ta có định hướng rõ ràng hơn về quyết định nên đầu tư cho công ty nào trong cùng một ngành để đem lại nguồn lợi tốt nhất.

Qua các kết quả từ phần mềm Orange trong các bài toán phân lớp và phân cụm. Nhóm nghiên cứu nhận thấy các kết quả ở bài toán về dự đoán cổ phiếu thì phương pháp

Tree là phương pháp phù hợp nhất khi cho ra các kết quả về Accuracy, F1- score, Precision và Recall đều là 99.4% cao hơn rất nhiều so với 2 mô hình còn lại từ 18.9% tới 26.8%. Bên cạnh đó, ma trận nhầm lẫn của Tree=1 nhỏ nhất trong cả ba phương pháp. Đối với bài toán phân tích khả năng nhận thêm vốn đầu tư, theo bảng đánh giá kết quả thì phương pháp Tree vẫn là phương pháp tốt nhất vì có các chỉ số Accuracy, F1- score, Precision và Recall lần lượt là 98.7%, 98.8%. 98.7% cao hơn phương pháp Logistic Regression và phương pháp SVM từ 10.9% đến 48.6%. Kết quả của ma trận nhầm lẫn, phương pháp Tree=0 nhỏ hơn 2 phương pháp còn lại. Ở bài toán cuối cùng, nhóm nghiên cứu thực hiện bài toán phân cụm, chúng ta thấy đối với phương pháp Hierarchical thì Ward trình bày và cho kết quả tốt nhất (phần lớn chỉ số silhouette đều trên 0.5 ở cụm đầu tiên, ở các cụm sau có ít chỉ số dưới 0). Và phương pháp K-means, lựa chọn phân thành 2 cụm là tốt nhất (các chỉ số Silhouette cao, ít chỉ số dưới 0).

Từ những số liệu phân tích được đã cho thấy rõ vấn đề tài chính ở các công ty hiện nay. Vô bọc tài chính bên ngoài của các công ty vô cùng hoàn hảo nhưng nếu như chúng ta không cẩn thận phân tích từng chỉ số tài chính một thì rất có thể dẫn tới sai lầm trong việc đầu tư. Bài nghiên cứu của nhóm cũng chứng minh được điều đó, bằng chứng là trong khi các chỉ số tài chính dù không được tốt nhưng lại cho ra kết quả cổ phiếu bị định giá cao rất nhiều. Việc một cổ phiếu được định giá cao ngoài thị trường tuy không hoàn toàn bị tác động bởi báo cáo tài chính mà còn phụ thuộc vào các yếu tố khác như: những dự án mới hay sản phẩm sắp ra mắt của các công ty có tiềm năng phát triển cao, hoặc những thành công trong sự quảng bá về các dự án tiềm năng thậm chí còn chưa bắt đầu nhưng đã bị các công ty quảng cáo quá đà với các nhà đầu tư dẫn đến việc họ kì vọng cao vào việc cổ phiếu của công ty đó sẽ mang lại lợi nhuận cao cho họ trong tương lai.

Nghiên cứu cũng đã chứng minh được việc cổ phiếu được định giá cao nhưng chưa chắc mang lại nguồn lợi lớn, ngược lại những cổ phiếu được định giá thấp hơn so với giá trị thực của chúng lại là một lựa chọn sáng suốt dành cho các nhà đầu tư, bởi sức khỏe tài chính ổn định là căn cứ vững chắc nhất để ta lựa chọn đầu tư, vì tính an toàn và lợi nhuận đều được đảm bảo hơn.

Cuối cùng, nhóm chúng em đã thực hiện được mục tiêu đã đề ra ban đầu là tổng hợp và phân tích báo cáo tài chính của 200 công ty ở Hoa Kỳ, xây dựng cơ sở giúp cho doanh nghiệp, các nhà đầu tư và bên cho vay có những chiến lược và mô hình kinh doanh, cũng như các quyết định đầu tư và cho vay phù hợp. Hơn thế nữa, việc phân tích báo cáo tài chính là một việc tất yếu của mỗi công ty trong quá trình hoạt động và phát triển của mình, mang lại những hiệu quả kinh doanh trong thực tiễn.

2. Những hạn chế

- Các thông tin thu thập của nghiên cứu có thể bị hạn chế do có quy mô nhỏ, chỉ phân tích tình hình tài chính của 200 công ty ở Hoa Kỳ.
- Quá trình tìm hiểu, tiếp cận và phân tích còn nhiều thiếu sót, chưa thể phân tích được một cách chi tiết và cụ thể.

- Sử dụng phương pháp chọn mẫu phi xác suất nên kết quả sẽ không được khách quan.

3. Hướng khắc phục

Từ những hạn chế nêu trên thì nhóm chúng em có đề ra những hướng khắc phục cụ thể như sau: mở rộng phạm vi nghiên cứu với đa quốc gia, đa dạng ngành nghề và ở nhiều lĩnh vực khác nhau, tìm hiểu và nghiên cứu kỹ càng về cách phân tích báo cáo tài chính cũng như đưa ra những phương hướng phù hợp cho các mục đích kinh tế khác nhau. Hơn thế nữa, chúng ta có thể sử dụng nhiều phương pháp nghiên cứu với độ chính xác cao để mang lại hiệu quả cao ở nhiều thời kỳ, nhiều giai đoạn để có những đánh giá rõ ràng và chính xác cho các doanh nghiệp. Đối với các nhà kinh doanh và đầu tư, phải nhận xét và đánh giá tình hình tài chính của công ty cũng như các phương pháp dự báo về các yếu tố quan trọng trước khi đưa ra các quyết định kinh tế hay hướng phát triển cho công ty. Họ nên đánh giá các lĩnh vực cũng như các ngành có tiềm năng phát triển trên thị trường và tập trung vào các công ty có khả năng tạo ra được lợi nhuận trong tương lai. Đồng thời nên có sự so sánh rõ ràng và đồng điệu, khi phải quyết định đầu tư hãy phân tích và so sánh số liệu tài chính của những công ty trong cùng một ngành để thấy rõ hơn sự khác biệt. Vì suy cho cùng mỗi ngành nghề đều có những hướng phát triển, kinh doanh khác nhau, vì thế khi so sánh ta phải so sánh các công ty cùng ngành để có cái nhìn khách quan nhất và đánh giá hoạt động của các công ty ấy một cách tốt nhất. Chính vì thế các dự báo cũng như kết quả của mô hình rất có ích cho các nhà kinh doanh, tạo cơ sở để đầu tư và phát triển.