

Hypothesis

Demographic Analysis (Phân tích theo nhân khẩu học):

- Dựa vào cột 'CODE_GENDER' ta có thể kiểm tra sự phân bố khách hàng theo giới tính.

Communication and Occupation Analysis (Kênh giao tiếp và nghề nghiệp):

- Điều tra xem các kênh liên lạc (FLAG_EMAIL) có liên quan đến việc trễ hạn sớm hay không.
- Kiểm tra ảnh hưởng của loại nghề nghiệp (OCCUPATION_TYPE) đến tỷ lệ nợ quá hạn.

Age and Employment Analysis (Phân tích độ tuổi và việc làm):

- Tính độ tuổi của người nộp đơn (DAYS_BIRTH) và đánh giá xem độ tuổi liên quan đến việc trễ hạn thế nào.
- Nghiên cứu ảnh hưởng của thời gian làm việc (DAYS_EMPLOYED) đến tình trạng nợ quá hạn.
- Xác định các xu hướng liên quan đến việc có điện thoại di động (FLAG_MOBIL), điện thoại cơ quan (FLAG_WORK_PHONE) và điện thoại cá nhân (FLAG_PHONE) khi nợ quá hạn.

Education and Family Analysis (Phân tích theo học vấn và gia đình của khách hàng):

- Dựa vào cột 'NAME_EDUCATION_TYPE' ta có thể phân tích trình độ học vấn ảnh hưởng thế nào đến tỷ lệ trả nợ quá hạn.
- Dựa vào cột 'NAME_FAMILY_STATUS' ta có thể phân tích tình trạng hôn nhân có thể ảnh hưởng đến tỷ lệ trả nợ quá hạn hay không.
- Dựa vào cột 'NAME_HOUSING_TYPE' ta có thể điều tra mối tương quan giữa loại nhà ở và tình trạng nợ quá hạn.
- Dựa vào cột 'CNT_CHILDREN' ta có thể phân tích xem việc khách hàng có con hoặc có bao nhiêu con có thể ảnh hưởng đến tỷ lệ trả nợ đúng hạn không.
- Dựa vào cột 'CNT_FAM_MEMBERS' Family Size Analysis: Nghiên cứu mối quan hệ giữa quy mô gia đình (CNT_FAM_MEMBERS) và tình trạng nợ quá hạn.

Property Ownership Analysis (Sở hữu tài sản):

- Dựa vào cột 'FLAG_OWN_CAR' ta có thể phân tích xem việc khách hàng sở hữu ô-tô có ảnh hưởng đến tỷ lệ trả nợ đúng hạn không.
- Dựa vào cột 'FLAG_OWN_REALTY' ta có thể phân tích xem việc khách hàng sở hữu bất động sản có ảnh hưởng đến tỷ lệ trả nợ đúng hạn không.

Income Analysis (Phân tích theo thu nhập khách hàng):

- Dựa vào cột 'AMT_INCOME_TOTAL' ta có thể xem xét mối quan hệ giữa thu nhập và tỷ lệ nợ quá hạn (thông qua...).
- Dựa vào cột 'NAME_INCOME_TYPE' ta có thể phân tích nguồn thu nhập với khả năng trả nợ sớm (vd: có nhiều nguồn thu nhập có thể có nhiều nguồn tiền vào nhiều thời điểm trong tháng).

Credit History Analysis (Phân tích lịch sử tín dụng):

Sử dụng tập dữ liệu [credit_record.csv] để hiểu lịch sử tín dụng (STATUS) của người vay theo thời gian. Phân tích các xu hướng quá hạn và nó phát triển như thế nào theo thời gian.

Historical Trends (Xu hướng tín dụng theo lịch sử):

Dựa vào cột 'MONTHS_BALANCE' ta có thể phân tích tỷ lệ quá hạn đã phát triển như thế nào theo thời gian của khách hàng.

Segment Analysis (Phân tích phân khúc):

Ta có thể phân loại nhóm người vay thành các phân khúc dựa trên [các thuộc tính khác nhau] và phân tích tỷ lệ quá hạn trong từng phân khúc đó.

Predictive Modeling (Mô hình dự đoán):

Ta có thể xây dựng predictive models bằng cách sử dụng [application_record.csv] để xác định các yếu tố có mối tương quan chặt chẽ với tình trạng nợ quá hạn (Có thể tìm hiểu trước về các kỹ thuật ML để dự đoán tình trạng nợ quá hạn).