**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG CÔNG NGHỆ - ĐẠI HỌC KINH TẾ QUỐC DÂN**

----------



**BÀI TẬP LỚN ỨNG DỤNG TRÍ TUỆ NHÂN TẠO**

**Đề tài:**

**Nghiên cứu và cài đặt chương trình phần mềm để thực hiện phân lớp khách hàng của ngân hàng bằng kỹ thuật Máy Vector hỗ trợ (SVM).**

**Người thực hiện : Nguyễn Minh Quang – 11225443**

**Phạm Minh Quân - 11225400**

**Môn học : Ứng dụng trí tuệ nhân tạo**

**Lớp học phần : TIHT1123(224)\_01**

**Giảng viên hướng dẫn : Ts.Phạm Thảo**

***Hà Nội, Năm 2025***

# MỤC LỤC

[MỤC LỤC 1](#_Toc197549503)

[DANH MỤC TỪ VIẾT TẮT 2](#_Toc197549504)

[LỜI MỞ ĐẦU 3](#_Toc197549505)

[I. TỔNG QUAN VỀ ĐỀ TÀI 4](#_Toc197549506)

[1.1.Phát biểu về đề tài: 4](#_Toc197549507)

[1.2. Mục đích, đối tượng và phạm vi nghiên cứu: 4](#_Toc197549508)

[1.3. Ý nghĩa khoa học và thực tiễn 5](#_Toc197549509)

[II. CƠ SỞ LÝ THUYẾT 6](#_Toc197549510)

[2.1.Đặt vấn đề 6](#_Toc197549511)

[2.2. Thuật toán SVM. 6](#_Toc197549512)

[2.2.1 Khái niệm 6](#_Toc197549513)

[2.2.2.Xây dựng bài toán: 7](#_Toc197549514)

[III.CÀI ĐẶT THỬ NGHIỆM 10](#_Toc197549515)

[3.1. Công nghệ áp dụng 10](#_Toc197549516)

[3.1.1. Python 10](#_Toc197549517)

[3.1.2. Các thư viện sử dụng. 10](#_Toc197549518)

[3.2. Giải thuật. 12](#_Toc197549519)

[3.3. Xử lý dữ liệu và chọn thuộc tính . 13](#_Toc197549520)

[3.4.Chia dữ liệu. 14](#_Toc197549521)

[3.5. Kết quả chương trình thử nghiệm. 15](#_Toc197549522)

[III.Triển khai giao diện 20](#_Toc197549523)

[IV. ĐÁNH GIÁ VÀ KẾT QUẢ NGHIÊN CỨU 22](#_Toc197549524)

[TÀI LIỆU THAM KHẢO 24](#_Toc197549525)

# DANH MỤC TỪ VIẾT TẮT

|  |  |  |
| --- | --- | --- |
| **STT** | **Từ viết tắt** | **Ý nghĩa** |
| **1** | **SVM** | **Support Vector Machine** |

**Link code:**

[**SVM-Bank/index.ipynb at main · minhquang2k4/SVM-Bank**](https://github.com/minhquang2k4/SVM-Bank/blob/main/index.ipynb)

# LỜI MỞ ĐẦU

Trong bối cảnh hiện đại, ngân hàng không chỉ là nơi thực hiện các giao dịch tài chính đơn thuần, mà còn đóng vai trò tư vấn và cung cấp các sản phẩm đầu tư – tiết kiệm phù hợp với từng đối tượng khách hàng. Một trong những thách thức lớn hiện nay của các ngân hàng là làm thế nào để xác định, phân loại được đâu là nhóm khách hàng có khả năng tham gia các chương trình tiết kiệm dài hạn, nhằm tối ưu hoạt động tư vấn và tiếp thị.

Trong khi dữ liệu về khách hàng ngày càng lớn và đa dạng, việc đánh giá thủ công là điều không còn phù hợp. Do đó, nhóm chúng em lựa chọn nghiên cứu đề tài “Phân lớp khách hàng ngân hàng bằng kỹ thuật Máy Vector Hỗ trợ (SVM)” nhằm ứng dụng trí tuệ nhân tạo vào việc xây dựng một mô hình hỗ trợ tự động trong phân loại khách hàng, từ đó giúp nhân viên ngân hàng dễ dàng biết được ai là người có tiềm năng tham gia gửi tiết kiệm dài hạn.

Đề tài không chỉ mang tính chất học thuật – nghiên cứu về kỹ thuật SVM, mà còn có ý nghĩa thực tiễn trong hoạt động tài chính ngân hàng. Việc áp dụng mô hình học máy giúp tăng tốc quy trình phân tích, nâng cao hiệu quả tiếp cận khách hàng và tiết kiệm chi phí nhân sự cho ngân hàng trong dài hạn.

Mặc dù thời gian và năng lực còn hạn chế, chúng em đã nỗ lực hoàn thành bài nghiên cứu với hy vọng sẽ nhận được những đóng góp quý báu từ thầy cô để cải thiện và phát triển sản phẩm này tốt hơn trong tương lai.

*Chúng em xin chân thành cảm ơn!*

# I. TỔNG QUAN VỀ ĐỀ TÀI

## 1.1.Phát biểu về đề tài:

Trong lĩnh vực tài chính – ngân hàng, việc xác định và phân loại khách hàng tiềm năng là một bước quan trọng nhằm nâng cao hiệu quả tiếp cận, tư vấn và triển khai các dịch vụ phù hợp. Đặc biệt, với các sản phẩm mang tính chất dài hạn như gửi tiết kiệm, việc dự đoán đúng nhóm khách hàng có khả năng tham gia là yếu tố then chốt giúp ngân hàng tối ưu nguồn lực và nâng cao tỷ lệ chuyển đổi.

Trước nhu cầu đó, nhóm lựa chọn đề tài “Phân lớp khách hàng ngân hàng bằng kỹ thuật Máy Vector Hỗ trợ (SVM)” với mục tiêu nghiên cứu và ứng dụng mô hình học máy nhằm tự động phân loại khách hàng thành hai nhóm chính: có khả năng và không có khả năng tham gia tiết kiệm dài hạn, dựa trên các đặc điểm nhân khẩu học và hành vi tài chính.

Kỹ thuật Support Vector Machine (SVM) là một trong những phương pháp phân lớp giám sát hiệu quả, nổi bật với khả năng xử lý tốt các bài toán phân lớp phi tuyến và hoạt động ổn định trong không gian dữ liệu nhiều chiều. Với đặc điểm này, SVM đã được ứng dụng rộng rãi trong nhiều lĩnh vực như tài chính, y tế, giáo dục… và đặc biệt phù hợp với bài toán phân lớp khách hàng trong ngân hàng hiện nay.

## 1.2. Mục đích, đối tượng và phạm vi nghiên cứu:

Mục đích của đề tài là nghiên cứu và xây dựng một mô hình phân lớp khách hàng ngân hàng dựa trên các đặc điểm cá nhân như giới tính, độ tuổi, khu vực sinh sống, tình trạng hôn nhân, thu nhập và một số yếu tố liên quan đến hành vi tài chính, nhằm dự đoán khả năng khách hàng có tham gia chương trình tiết kiệm dài hạn hay không.

Đối tượng nghiên cứu là tập dữ liệu khách hàng do ngân hàng cung cấp (Bank-data), trong đó mỗi khách hàng được gắn nhãn thể hiện khả năng có/không tham gia gửi tiết kiệm (thông qua thuộc tính "pep").

Phạm vi nghiên cứu tập trung vào việc sử dụng mô hình Support Vector Machine (SVM) để thực hiện phân lớp, từ đó đánh giá độ chính xác, khả năng áp dụng mô hình vào thực tế, đồng thời đề xuất các hướng cải tiến trong tương lai như mở rộng dữ liệu, tối ưu hóa thuật toán hoặc tích hợp vào hệ thống CRM nội bộ của ngân hàng.

## 1.3. Ý nghĩa khoa học và thực tiễn

Về mặt khoa học, đề tài giúp người học nắm vững kiến thức lý thuyết và cách triển khai thực tế của mô hình Support Vector Machine (SVM) – một trong những kỹ thuật phân lớp mạnh mẽ và phổ biến nhất trong lĩnh vực học máy hiện nay. Việc tìm hiểu, cài đặt và đánh giá hiệu quả mô hình trên dữ liệu thực tế không chỉ củng cố kiến thức về thuật toán mà còn rèn luyện tư duy xử lý dữ liệu, phân tích kết quả và tối ưu mô hình.

Về mặt thực tiễn, mô hình phân lớp SVM có thể được ứng dụng hiệu quả trong lĩnh vực ngân hàng, cụ thể là hỗ trợ tự động nhận diện nhóm khách hàng tiềm năng có khả năng tham gia chương trình tiết kiệm dài hạn. Đây là một nhu cầu thiết thực giúp ngân hàng nâng cao hiệu quả tiếp thị, tối ưu hóa nguồn lực tư vấn và cải thiện trải nghiệm khách hàng. Trong bối cảnh dữ liệu khách hàng ngày càng lớn và đa dạng, việc ứng dụng trí tuệ nhân tạo để hỗ trợ ra quyết định là xu hướng tất yếu trong chuyển đổi số của ngành ngân hàng.

Ngoài phần Lời mở đầu và Tài liệu tham khảo và Phụ lục thì nghiên cứu gồm 4 chương chính:

*Chương 1: Tổng quan về đề tài*

*Chương 2: Cơ sở lý thuyết.*

*Chương 3: Cài đặt thử nghiệm*

*Chương 4: Đánh giá kết quả nghiên cứu*

# II. CƠ SỞ LÝ THUYẾT

## 2.1.Đặt vấn đề

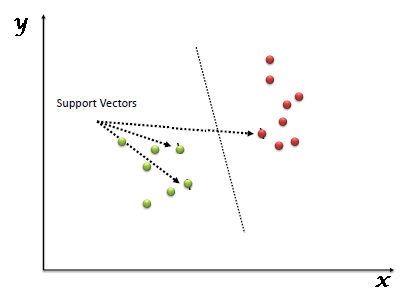
Phân loại thống kê là một nhiệm vụ phổ biến trong ứng dụng trí tuệ nhân tạo. Trong mô hình học có giám sát, thuật toán được cho trước một số điểm dữ liệu cùng với nhãn của chúng thuộc một trong hai lớp cho trước. Mục tiêu của thuật toán là xác định xem một điểm dữ liệu mới sẽ được thuộc về lớp nào. Mỗi điểm dữ liệu được biểu diễn dưới dạng một vector p-chiều, và ta muốn biết liệu có thể chia tách hai lớp dữ liệu bằng một siêu phẳng p − 1 chiều. Đây gọi là phân loại tuyến tính. Có nhiều siêu phẳng có thể phân loại được dữ liệu. Một lựa chọn hợp lý trong chúng là siêu phẳng có lề lớn nhất giữa hai lớp

.

## 2.2. Thuật toán SVM.

### 2.2.1 Khái niệm

Máy vector hỗ trợ (SVM - Support Vector Machine ) là một thuật toán giám sát giúp tìm ra một siêu phẳng phân cách tối ưu để có thể phân chia dữ liệu tuyến tính ra làm hai lớp khác nhau, được sử dụng để phân loại, hồi qui và phát hiện điểm dữ liệu bất thường. Tuy nhiên nó được sử dụng chủ yếu cho việc phân loại.

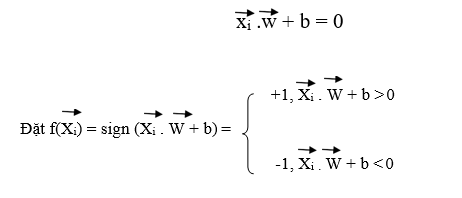
Support Vectors hiểu một cách đơn giản là các đối tượng trên đồ thị tọa độ quan sát, Support Vector Machine là một biên giới để chia hai lớp tốt nhất.

### 2.2.2.Xây dựng bài toán:

SVM thực chất là một bài toán tối ưu, mục tiêu của thuật toán này là tìm được một không gian F và siêu phẳng quyết định f trên F sao cho sai số phân loại là thấp nhất.

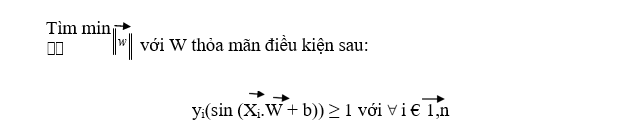
Cho tập mẫu (x1, y1), (x2, y2), … (xf, yf )} với xi ∈ Rn , thuộc vào hai lớp nhãn: yi ∈ {-1,1} là nhãn lớp tương ứng của các xi (-1 biểu thị lớp I, 1 biểu thị lớp II).

Ta có, phương trình siêu phẳng chứa vectơ xi trong không gian:





Như vậy, f(Xi) biểu diễn sự phân lớp của Xi vào hai lớp như đã nêu. Ta nói yi= +1 nếu Xi € lớp I và yi = -1 nếu Xi € lớp II . Khi đó, để có siêu phẳng f ta sẽ phải giải bài toán sau:



Bài toán SVM có thể giải bằng kỹ thuật sử dụng toán tử Lagrange để biến đổi về thành dạng đẳng thức. Một điểm thú vị của SVM là mặt phẳng quyết định chỉ phụ thuộc vào các Support Vector và nó có khoảng cách đến mặt phẳng quyết định là . Cho dù các điểm khác bị xóa thì thuật toán vẫn cho kết quả giống như ban đầu. Đây chính là điểm nổi bật của phương pháp SVM so với các phương pháp khác vì tất cả các dữ liệu trong tập huấn luyện đều được dùng để tối ưu hóa kết quả.

***TÓM LẠI:*** trong trường hợp nhị phân phân tách tuyến tính, việc phân llớp được thực hiện qua hàm quyết định *f(x) = sign(<w.x> + b),* hàm này thu được bằng việc thay đổi vectơ chuẩn *w*, đây là vectơ để cực đại hóa viền chức năng

Việc mở rộng SVM để phân đa lớp hiện nay vẫn đang được đầu tư nghiên cứu. Có một phương pháp tiếp cận để giải quyết vấn để này là xây dựng và kết hợp nhiều bộ phân lớp nhị phân SVM (Chẳng hạn: trong quá trình luyện với SVM, bài toán phân m lớp có thể được biến đổi thành bài toán phân 2\*m lớp, khi đó trong mỗi hai lớp, hàm quyết định sẽ được xác định cho khả năng tổng quát hóa tối đa). Trong phương pháp này có thể đề cập tới hai cách là *một-đổi-một, một-đối-tất cả*

2.2.3.Ví dụ minh họa:

Ứng dụng Support Vector Machine trong bài toán phân loại hoa

* Tập dữ liệu Iris Flowers

Tập dữ liệu này có sẵn trong thư viện skit-learn gồm 50 mẫu về 3 loài hoa khác nhau của họ Iris là (Iris setosa, Iris virginica và Iris versicolor).

Với mỗi một mẫu hoa này có bốn thuộc tính là chiều dài và chiều rộng của đài hoa và cánh hoa với đơn vị centimet.

* Sau quá trình huấn luyện ta thu được đồ thị :



Việc phân lớp chia mặt phẳng tọa độ thành các phần khác nhau. Từ đó, khi có một điểm dữ liệu mới có thể dựa vào tọa độ của chúng để phán đoán xem nó thuộc lớp nào.

# III.CÀI ĐẶT THỬ NGHIỆM

Trong chương này, nhóm tiến hành cài đặt mô hình SVM trên tập dữ liệu khách hàng của ngân hàng, với mục tiêu kiểm tra khả năng mô hình trong việc phân loại khách hàng thành hai nhóm: có khả năng và không có khả năng tham gia tiết kiệm dài hạn. Việc phân loại này có thể giúp các nhân viên ngân hàng nhanh chóng xác định đúng nhóm khách hàng tiềm năng, hỗ trợ cho hoạt động tư vấn sản phẩm hiệu quả hơn.

## 3.1. Công nghệ áp dụng

Đối với phần cài đặt thử nghiệm này, nhóm đã sử dụng ngôn ngữ lập trình Python3 cùng với các thư viện liên quan đến việc huấn luyện mô hình như: Matplotlib, Scikit-learn, Pandas, Numpy.

### 3.1.1. Python

Python là một ngôn ngữ lập trình thông dịch (interpreted), hướng đối tượng (object-oriented), và là một ngôn ngữ bậc cao (high-level) ngữ nghĩa động (dynamic semantics). Python hỗ trợ các module và gói (packages), khuyến khích chương trình module hóa và tái sử dụng mã. Trình thông dịch Python và thư viện chuẩn mở rộng có sẵn dưới dạng mã nguồn hoặc dạng nhị phân miễn phí cho tất cả các nền tảng chính và có thể được phân phối tự do.

Python cung cấp mã ngắn gọn và dễ đọc. Trong khi các thuật toán của AI và Machine learning là hết sức phức tạp, nhưng sự đơn giản của Python cho phép các nhà phát triển tạo nên các hệ thống rất chính xác và đáng tin cậy. Ngoài ra Python còn có rất nhiều thư viện hỗ trợ cho việc huấn luyện mô hình.

### 3.1.2. Các thư viện sử dụng.

- Numpy: là một thư viện toán học phổ biến và mạnh mẽ của Python. Cho phép làm việc hiệu quả với ma trận và mảng, đặc biệt là dữ liệu ma trận và mảng lớn với tốc độ xử lý nhanh hơn nhiều lần khi chỉ sử dụng “core Python” đơn thuần.

- Scikit-learn (Sklearn) là thư viện mạnh mẽ nhất dành cho các thuật toán ứng dụng trí tuệ nhân tạo được viết trên ngôn ngữ Python. Thư viện cung cấp một tập các công cụ xử lý các bài toán machine learning và statistical modeling

- Matplotlib là một thư viện vẽ đồ thị rất mạnh mẽ hữu ích cho những người làm việc với Python và NumPy.

- Pandas là một thư viện để thực hiện phân tích và thao tác dữ liệu; bất kỳ loại xử lý, phân tích, lọc và tổng hợp dữ liệu nào. Pandas là một trong những công cụ quan trọng trong việc hỗ trợ, xử lý và phân tích dữ liệu với mã nguồn mở nhanh, mạnh, linh hoạt và dễ sử dụng,

## 3.2. Giải thuật.

**Diagram, engineering drawing

Description automatically generated**

Các bước thực hiện quá trình huấn luyện và xác định được thể hiện trong dữ liệu từ Bank-data ban đầu được xử lý , chọn ra các thuộc tính cần thiết cho huấn luyện rồi được tác thành 2 tập là tập train và tập test. Trong đó, Bank-data là tập dữ liệu được cho ban đầu (bao gồm 600 khách hàng ), tập dùng để training (huấn luyện), tập dùng để test (kiểm tra). Trong quá trình thực nghiệm, do số mẫu ít nên tập huấn luyện và tập kiểm tra được lấy ngẫu nhiên từ tập Bank-data với tỷ lệ lần lượt là 80% và 20% . Sau quá trình huấn luyện, ta cho tiến hành dự đoán kết quả dựa theo tập kiểm tra thu được kết quả dự đoán rồi so sánh kết quả dự đoán này với kết quả của tập test.

## 3.3. Xử lý dữ liệu và chọn thuộc tính .

A picture containing diagram

Description automatically generated

Xét tập dữ liệu Bank-data :Table, Excel

Description automatically generated

*Hình 3.2.1. Dữ liệu của Bank-data (9 dòng đầu)*

Tập dữ liệu ban đầu gồm 11 thuộc tính trong đó thuộc tính “ID ” không sử dụng trong huấn luyện mô hình, thuộc tính “pep” là kết quả và các thuộc tính còn lại được sử dụng trong việc đánh giá mô hình và kết quả dự đoán.

Xét các thuộc tính có dữ liệu dạng văn bản, để dàng cho việc tính toán nhóm đã chuyển các dữ liệu này về dạng số thực. Việc thay đổi được thực hiện theo quy tắc sau :Table

Description automatically generated

Hình 3.2.2. Chuyển đổi dữ liệu

* Thuộc tính ‘Sex’ có 2 giá trị : ‘Female’ và ‘Male’ được gán với giá trị trương ứng lần lượt là 0 và 1
* Thuộc tính ‘Region’ có 4 giá trị ‘Rural’, ‘Inner\_City’, ‘Suburban’, ‘Town’ được gán với giá trị trương ứng lần lượt là 0, 1, 2 và 3
* Các thuộc tính ‘Married’, ‘Car’, ‘Save\_Act’, ‘Current\_Act’, ‘Mortgage’, ‘Pep’ đều có 2 giá trị : ‘YES’, ‘NO’ được gán với giá trị trương ứng lần lượt là 0 và 1

Sau khi chuyển đổi ta được tập dữ liệu sau khi xử lý :Chart

Description automatically generated

*Hình 3.2.3. Dữ liệu của Bank-data sau khi xử lý (9 dòng đầu)*

## 3.4.Chia dữ liệu.

Sau khi xử lý dữ liệu về dạng có thể xửa lý được, nhóm tiến hành chia ngẫu nhiên tập dữ liệu này thành 2 phần là tập huấn luyện và tập kiểm tra. Tập kiểm tra chiếm 20% dữ liệu lấy vào trương đương với 120 khách hàng. Bởi vì dữ liệu cho việc huấn luyện và kiểm tra được lấy ngẫu nhiên, kết quả kiểm tra thu thập được sau mỗi lần đều có sự khác nhau.

## 3.5. Kết quả chương trình thử nghiệm.

Nhóm xin được trình bày ra kết quả của 5 lần chạy thử nghiệm để có góc nhìn tổng quan:

Lần 1: Chart, treemap chart

Description automatically generated

Kết quả:

Dự đoán đúng 70/120 ~ 58% kết quả so với tập kiểm tra trong đó:

* 16/56 ~ 29% kết quả dự đoán đúng đối với đáp án “Yes” tức giá trị 0
* 54/64 ~ 84 % kết quả dự đoán đúng đối với đáp án “No” tức giá trị 1
* Lần 2:Chart

  Description automatically generated

Kết quả:

Dự đoán đúng 69/120 ~ 58% kết quả so với tập kiểm tra trong đó:

* 14/51 ~ 27% kết quả dự đoán đúng đối với đáp án “Yes” tức giá trị 0
* 55/65 ~ 84 % kết quả dự đoán đúng đối với đáp án “No” tức giá trị 1
* Lần 3: Chart

  Description automatically generated

Kết quả:

Dự đoán đúng 66/120 ~ 56% kết quả so với tập kiểm tra trong đó:

* 9/57 ~ 16% kết quả dự đoán đúng đối với đáp án “Yes” tức giá trị 0
* 57/63 ~ 90% kết quả dự đoán đúng đối với đáp án “No” tức giá trị 1
* Lần 4Chart, treemap chart

  Description automatically generated

Kết quả:

Dự đoán đúng 79/120 ~ 67% kết quả so với tập kiểm tra trong đó:

* 20/48 ~ 42% kết quả dự đoán đúng đối với đáp án “Yes” tức giá trị 0
* 59/72 ~ 82% kết quả dự đoán đúng đối với đáp án “No” tức giá trị 1
* **Lần 5**Chart

  Description automatically generated

Kết quả:

Dự đoán đúng 66/120 ~ 56% kết quả so với tập kiểm tra trong đó:

* 15/54 ~ 28% kết quả dự đoán đúng đối với đáp án “Yes” tức giá trị 0
* 51/66 ~ 77% kết quả dự đoán đúng đối với đáp án “No” tức giá trị 1

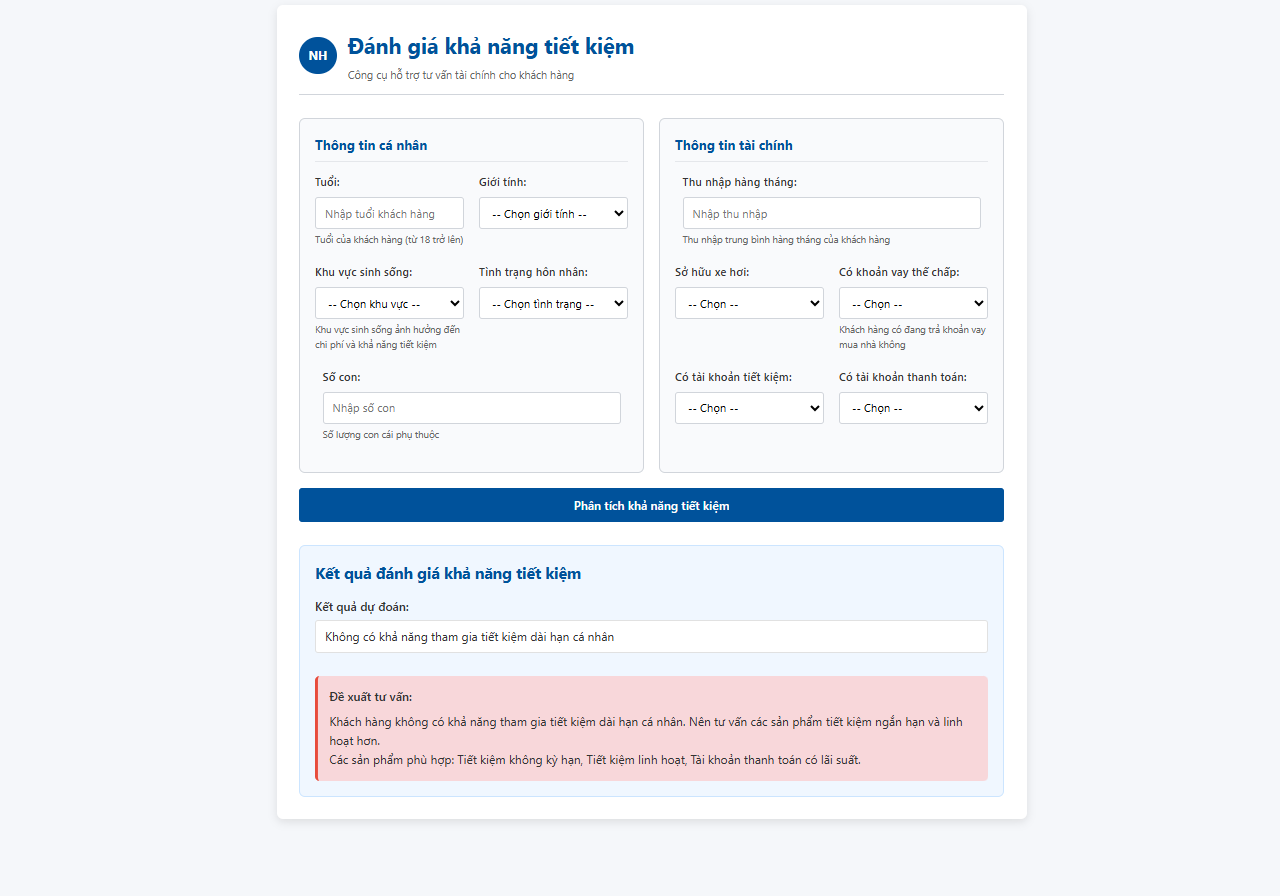
# III.Triển khai giao diện

A screenshot of a computer

AI-generated content may be incorrect.

*Giao diện chính*

Hình trên là giao diện chính của mô hình dự đoán. Giao diện này được thiết kế để người dùng nhập các thông tin khách hàng, bao gồm các thuộc tính như giới tính, độ tuổi, khu vực sinh sống, tình trạng hôn nhân, thu nhập, và các yếu tố liên quan đến hành vi tài chính. Các trường nhập liệu được bố trí rõ ràng, trực quan, đảm bảo người dùng có thể nhập dữ liệu chính xác và thuận tiện.



*Kết quả sau khi phân tích khả năng tiết kiệm*

Kết quả sau quá trình mô hình tiến hành phân tích dự đoán được hiển thị. Với mỗi kết quả dự đoán sẽ có đề xuất tư vấn tương ứng với kết quả dự đoán đó. Hình trên hiển thị kết quả dự đoán khách hàng không có khả năng tham gia tiết kiệm dài hạn cá nhân cùng với đó là đề xuất tư vấn nên tư vấn các sản phẩm tiết kiệm ngắn hạn và linh hoạt. Ngoài ra mô hình còn đề xuất các sản phẩm phù hợp với đối tượng khách hàng.

**A screenshot of a computer

AI-generated content may be incorrect.**

*Kết quả sau khi phân tích khả năng tiết kiệm*

Hình trên hiển thị kết quả dự đoán của một khách hàng có khả năng tham gia tiết kiệm dài hạn. Tương tự như trên, từ kết quả dự đoán sẽ có đề xuất tư vấn tương ứng với kết quả, phù hợp với từng đối tượng khách hàng.

# IV. ĐÁNH GIÁ VÀ KẾT QUẢ NGHIÊN CỨU

Qua quá trình nghiên cứu và thực nghiệm, nhóm đã bước đầu xây dựng thành công mô hình phân lớp khách hàng sử dụng thuật toán Máy Vector Hỗ trợ (SVM). Mô hình cho kết quả khả quan, đặc biệt là trong việc phân biệt nhóm khách hàng không có khả năng tham gia tiết kiệm dài hạn với độ chính xác tương đối cao. Tuy nhiên, độ chính xác với nhóm khách hàng có khả năng tham gia còn hạn chế do sự mất cân bằng trong dữ liệu đầu vào.

Kết quả này đã mở ra hướng tiếp cận rõ ràng cho việc ứng dụng trí tuệ nhân tạo vào hệ thống hỗ trợ ra quyết định trong lĩnh vực ngân hàng. Cụ thể, mô hình có thể tích hợp vào hệ thống quản lý khách hàng (CRM) nhằm giúp nhân viên tư vấn ưu tiên tiếp cận nhóm khách hàng tiềm năng, từ đó tối ưu hóa nguồn lực và nâng cao tỷ lệ chuyển đổi trong các chiến dịch tiết kiệm.

Mặc dù thời gian thực hiện có giới hạn, nhóm đã hoàn thành được các mục tiêu cốt lõi như:

* Tìm hiểu và trình bày chi tiết về phương pháp SVM – một kỹ thuật phân lớp mạnh mẽ và phổ biến trong học máy.
* Cài đặt thành công chương trình huấn luyện và dự đoán bằng ngôn ngữ Python, sử dụng các thư viện chuyên dụng như Scikit-learn, Pandas, NumPy.
* Thực hiện xử lý dữ liệu đầu vào, trích đặc trưng và đánh giá hiệu quả mô hình trên tập dữ liệu thực tế.

Trong tương lai, nhóm kỳ vọng sẽ tiếp tục cải tiến mô hình bằng cách:

* Bổ sung thêm dữ liệu đa dạng và cân bằng hơn.
* Kết hợp thêm các thuật toán học máy nâng cao như Random Forest, XGBoost hoặc mạng nơ-ron.
* Tối ưu hóa quá trình tiền xử lý dữ liệu và đánh giá kết quả theo nhiều chỉ số khác nhau (precision, recall, F1-score…).

Từ đó, đề tài có thể được phát triển thành một hệ thống hỗ trợ ra quyết định hoàn chỉnh, sẵn sàng ứng dụng trong môi trường thực tế của các ngân hàng thương mại.

# TÀI LIỆU THAM KHẢO

[1] Việt Nam Biz, “Máy vector hỗ trợ (Support Vector Machine – SVM) là gì?”, *Vietnambiz.vn*, 26-Feb-2020. [Trực tuyến]. Có tại: <https://vietnambiz.vn/may-vector-ho-tro-support-vector-machine-svm-la-gi-20200226223210903.html>

[2] Wikipedia, “Máy vectơ hỗ trợ”, *Wikipedia tiếng Việt*. [Trực tuyến]. Có tại: <https://vi.wikipedia.org/wiki/M%C3%A1y_vect%C6%A1_h%E1%BB%97_tr%E1%BB%A3>

[3] Viblo, “Giới thiệu về Support Vector Machine (SVM)”, *Viblo.asia*. [Trực tuyến]. Có tại: <https://viblo.asia/p/gioi-thieu-ve-support-vector-machine-svm-6J3ZgPVElmB>

[4] Viblo, “Ứng dụng Support Vector Machine trong bài toán phân loại hoa”, *Viblo.asia*. [Trực tuyến]. Có tại: <https://viblo.asia/p/ung-dung-support-vector-machine-trong-bai-toan-phan-loai-hoa-PdbGnLXBkyA>