|  |
| --- |
| **BỘ GIÁO DỤC VÀ ĐÀO TẠO**  **TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH**  **KHOA CÔNG NGHỆ THÔNG TIN**  Description: C:\Documents and Settings\Administrator\Desktop\logo dai hoc_khong nen.png  **TIỂU LUẬN MÔN HỌC**  **ĐỒ ÁN CHUYÊN NGÀNH KHOA HỌC DỮ LIỆU**  **PHÂN TÍCH DỮ LIỆU VỀ TÀI CHÍNH NGÂN HÀNG VCB**  Giảng viên giảng dạy: Th.S VƯƠNG XUÂN CHÍ  Sinh viên thực hiện: LÊ XUÂN TRỌNG  MSSV : 2100009483  Chuyên ngành : KHOA HỌC DỮ LIỆU  Môn học : ĐỒ ÁN CHUYÊN NGÀNH KHOA HỌC DỮ LIỆU  Khóa : 2021    **Tp.HCM, tháng 09 Năm 2024** |
| **BỘ GIÁO DỤC VÀ ĐÀO TẠO**  **TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH**  **KHOA CÔNG NGHỆ THÔNG TIN**  Description: C:\Documents and Settings\Administrator\Desktop\logo dai hoc_khong nen.png  **TIỂU LUẬN MÔN HỌC**  **ĐỒ ÁN CHUYÊN NGÀNH KHOA HỌC DỮ LIỆU**  **PHÂN TÍCH DỮ LIỆU VỀ TÀI CHÍNH NGÂN HÀNG VCB**  Giảng viên giảng dạy:Th.S VƯƠNG XUÂN CHÍ  Sinh viên thực hiện: LÊ XUÂN TRỌNG  MSSV : 2100009483  Chuyên ngành : KHOA HỌC DỮ LIỆU  Môn học : ĐỒ ÁN CHUYÊN NGÀNH KHOA HỌC DỮ LIỆU  Khóa : 2021    **Tp.HCM, tháng 09 Năm 2024** |

|  |  |
| --- | --- |
| Trường Đại học Nguyễn Tất Thành  **Khoa Công Nghệ Thông Tin**  🙜  🙜  🙝  🙝 | CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM  **Độc lập – Tự do – Hạnh phúc**  🙜 🙜 🙝 🙝 |

**NHIỆM VỤ ĐỒ ÁN CHUYÊN NGÀNH**

Họ và tên: **LÊ XUÂN TRỌNG** MSSV: **2100009483**

Chuyên ngành: **Khoa học dữ liệu** Lớp: **21DTH2C**

Email: **xuantrongle987@gmail.com** SĐT: **0366394982**

Tên đề tài: Phân tích dữ liệu về tài chính ngân hàng VCB

Giảng viên hướng dẫn: **ThS.Vương Xuân Chí**

Thời gian thực hiện:  **24 /06 /2024 đến 05/ 09 /2024**

**MÔ TẢ ĐỀ TÀI:**

Xem xét các chỉ số lợi nhuận như doanh thu, lợi nhuận trước thuế, lợi nhuận ròng và biên lợi nhuận. Xem xét các khoản nợ của ngân hàng, bao gồm nợ vay từ khách hàng và nợ vay từ các tổ chức khác. Thông tin về số lượng và phân loại khách hàng, Trực quan hóa, (xây dựng thực nghiệm)

**NỘI DUNG VÀ PHƯƠNG PHÁP:**

* Giới thiệu về Phân tích dữ liệu tài chính ngân hàng VCB
* Mô tả mô hình phân tích dữ liệu tài chính
* Các ứng dụng trong công nghệ cũng như trong cuộc sống
* Cài đặt mô phỏng,
* Kết luận

**YÊU CẦU:**

* Có kiến thức, đam mê, hiểu biết về Machine Learning, Deep Learning. …Đọc hiểu tài liệu tiếng Anh, kỹ năng trình bày văn bản trên máy tính tốt.
* Có tác phong làm việc chăm chỉ, tinh thần trách nhiệm cao, có khả năng làm việc độc lập hoặc làm việc trong nhóm tốt.

**Nội dung và yêu cầu đã được thông qua Bộ môn.**

*TP.HCM, ngày 10 tháng 6 năm 2024*

|  |  |
| --- | --- |
| **Q. TRƯỞNG BỘ MÔN**  *(Ký và ghi rõ họ tên)* | **GIÁO VIÊN HƯỚNG DẪN**  *(Ký và ghi rõ họ tên)* |

**ThS. Vương Xuân Chí ThS. Vương Xuân Chí**

# LỜI MỞ ĐẦU

Trong môi trường cạnh tranh khốc liệt của ngành tài chính ngân hàng hiện nay, phân tích dữ liệu tài chính đóng vai trò quan trọng trong việc ra quyết định và quản lý rủi ro. Ngân hàng Vietcombank (VCB), một trong những ngân hàng hàng đầu tại Việt Nam, đang đối diện với áp lực từ sự biến động của thị trường và yêu cầu ngày càng cao về độ chính xác trong các hoạt động tài chính.

Nghiên cứu này tập trung vào việc áp dụng các phương pháp phân tích dữ liệu để đánh giá các chỉ số tài chính của Ngân hàng VCB. Chúng tôi sẽ khai thác dữ liệu từ báo cáo tài chính, bảng cân đối kế toán, và các chỉ số hoạt động chính nhằm tìm hiểu các xu hướng tài chính, đánh giá hiệu quả hoạt động và dự đoán các rủi ro tiềm ẩn. Mục tiêu là cung cấp cái nhìn sâu sắc về tình hình tài chính hiện tại của VCB và hỗ trợ việc phát triển các chiến lược tài chính cùng phương pháp quản lý rủi ro hiệu quả.

Cuối cùng, nghiên cứu này nhắm đến việc xây dựng một mô hình phân tích dữ liệu chính xác và đáng tin cậy, nhằm giúp ngân hàng đưa ra quyết định thông minh và nhanh chóng, từ đó nâng cao hiệu quả hoạt động và khả năng cạnh tranh trong thị trường tài chính.

**LỜI CẢM ƠN**

Nếu như phải gửi một lời cảm ơn dành cho các thầy cô đã trợ giúp cho mình thì cho em xin phép mượn câu nói của ngài William a Warrrd, ông từng nói rằng:

“Một người thầy trung bình chỉ biết nói

Một người thầy giỏi biết giải thích

Một người thầy chúng biết minh họa

Một người thầy vĩ đại biết truyền cảm hứng”

Nếu phải dành lời cảm ơn em muốn mượn câu nói này để dành cho các thầy cô phụ trách, cũng như là thầy Vương Xuân Chí giảng viên phụ trách môn đồ án chuyên ngành Khoa Học Dữ Liệu đã giúp cũng như hỗ trợ chúng em hoàn thành được đồ án về chủ đề phân tích dữ liệu về tài chính ngân hàng VCB. Cũng từ đồ án này em nhận ra mình cần phải có được tội lỗi phàm ăn để khơi lên cơn đói khát về kiến thức của mình từ đó ăn bớt những hạn chế của mình trong con sông kiến thức cũng như tránh xa các tội lỗi kiêu ngạo, lười biếng và đố kị để làm tròn nên những kinh nghiệm mà mình cần phải học hỏi từ mọi người để bước gần hơn tới thành công. Em cũng mong nhận được những lời gớp ý giúp em cải thiện cũng hoàn thiện đồ án hơn.

Cuối cùng em xin kính chúc quí thầy cô có một ngày tốt lạnh.

Em chân thành cảm ơn

Sinh viên thực hiện

LÊ XUÂN TRỌNG

|  |  |
| --- | --- |
| TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH  **TRUNG TÂM KHẢO THÍ** | **KỲ THI KẾT THÚC HỌC PHẦN**  **HỌC KỲ II NĂM HỌC 2023 - 2024** |

**PHIẾU CHẤM THI TIỂU LUẬN/ĐỒ ÁN**

BM-ChT-11

Môn thi: Đồ án chuyên ngành khoa học dữ liệu Lớp học phần:21DTH2C

Nhóm sinh viên thực hiện:

1. Lê Xuân Trọng Tham gia đóng góp:100%

2. Tham gia đóng góp:

3. Tham gia đóng góp:

4. Tham gia đóng góp:

5. Tham gia đóng góp:

6. Tham gia đóng góp:

7. Tham gia đóng góp:

8. Tham gia đóng góp:

Ngày thi: 11/09/2024 Phòng thi: L.506

Đề tài tiểu luận/báo cáo của sinh viên : Phân tích dữ liệu về tài chính ngân hàng VCB

|  |  |  |  |
| --- | --- | --- | --- |
| **Tiêu chí (theo CĐR HP)** | **Đánh giá của GV** | **Điểm tối đa** | **Điểm đạt được** |
| Cấu trúc của báo cáo | Gồm 3 chương | 1 |  |
| Nội dung |  |  |  |
| * Các nội dung thành phần | Chương 2 và 3 | 5 |  |
| * Lập luận | Chương 1 | 2 |  |
| * Kết luận |  | 1 |  |
| Trình bày | Theo chuẩn format luận văn font chữ 13, canh trái, phải,… | 1 |  |
| **TỔNG ĐIỂM** |  | **10** |  |

|  |  |
| --- | --- |
|  | **Giảng viên chấm thi**  *(ký, ghi rõ họ tên)* |

**NHẬN XÉT CỦA GIẢNG VIÊN GIẢNG DẠY**

*Tp.HCM, Ngày 28 tháng 12 năm 2022*

**Giảng viên giảng dạy**

(Ký tên và ghi rõ họ tên)

**MỤC LỤC**

[LỜI MỞ ĐẦU iv](#_Toc15110)

[LỜI CẢM ƠN v](#_Toc6157)

[CHƯƠNG 1: GIỚI THIỆU 1](#_Toc22830)

[1.1 Giới thiệu đề tài. 1](#_Toc16537)

[1.2.Lý do chọn đề tài 2](#_Toc26314)

[1.3.Mục tiêu đề tài. 3](#_Toc29967)

[1.4.Phương pháp đề tài. 4](#_Toc18831)

[1.5.Đối tượng và phạm vi nghiên cứu. 5](#_Toc8276)

[1.5.1. Đối Tượng Nghiên Cứu 5](#_Toc26795)

[1.5.2. Phạm Vi Nghiên Cứu 6](#_Toc3430)

[CHƯƠNG 2 ỨNG DỤNG VÀ THUẬT TOÁN 7](#_Toc15075)

[2.1 Giới thiệu bài toán. 7](#_Toc24720)

[2.2 Mô tả thuật toán. 8](#_Toc8588)

[2.2.1. Thuật toán Disision Tree 8](#_Toc21311)

[2.2.2. Thuật toán Random Forest 8](#_Toc33)

[2.2.3. Thuật toán K-Nearest Neighbors 9](#_Toc5052)

[2.3. Xây dựng bộ dữ liệu. 9](#_Toc10647)

[2.3.1. Bộ dữ liệu gồm các cột  9](#_Toc22447)

[2.3.2. Tiền xử lý dữ liệu  10](#_Toc6491)

[2.4. Áp dụng thuật toán vào bài toán. 10](#_Toc24497)

[2.5. Thực nghiệm với thư viện. 11](#_Toc23825)

[CHƯƠNG 3 XÂY DỰNG ỨNG DỤNG NGÔN NGỮ PYTHON 13](#_Toc23673)

[3.1 Xem xét các chỉ số lợi nhuận 13](#_Toc32152)

[3.1.1. Doanh Thu 13](#_Toc27240)

[3.1.2. Lợi nhuận trước thuế 14](#_Toc26845)

[3.1.3. Lợi nhuận ròng 15](#_Toc16001)

[3.1.4. Biên lợi nhuận 15](#_Toc11130)

[3.1.5. Tổng nợ cho khách hàng vay 16](#_Toc14351)

[3.1.6. Tổng nợ vay từ các tổ chức khác 17](#_Toc66)

[3.1.7. Số lượng và phân loại khách hàng 18](#_Toc25732)

[3.2 Xây dựng ứng dụng và giải thích 20](#_Toc11550)

[3.2.1. Random Forest 20](#_Toc12124)

[3.2.2. K-Nearest Neighbors 21](#_Toc8174)

[3.2.3. Decision Tree Classifier 21](#_Toc1543)

[3.2.4. So sanh các phương pháp dự đoán 22](#_Toc4932)

[3.3. Dự đoán các chỉ số cho tương lai 22](#_Toc14370)

[3.3.1. Dự đoán doanh thu 22](#_Toc570)

[3.3.2. Dự đoán Lợi nhuận trước thuế 22](#_Toc24205)

[3.3.3. Dự đoán lợi nhuận ròng 23](#_Toc11184)

[3.3.4. Dự đoán biên lợi nhuận 23](#_Toc10274)

[3.3.5. Dự đoán Nợ cho khách hàng vay 24](#_Toc10415)

[3.3.6. Dự đoán Nợ vay từ các tổ chức khác 24](#_Toc8461)

[3.4. Một số ảnh trực phân hóa được phân tích trong dự án 24](#_Toc30416)

[3.4.1. Trực quan hoá doanh thu(Revenue) 25](#_Toc29219)

[3.4.2. Trực quan hóa Lợi nhuận trước thuế 25](#_Toc7881)

[3.4.3. Trực quan hóa Lợi nhuận ròng theo năm 26](#_Toc16291)

[3.4.4. Trực quan hóa Biên lợi nhuận theo năm 27](#_Toc8350)

[3.4.5. Trực quan hóa Các khoản vay của ngân hàng 28](#_Toc19269)

[3.4.6. Trực quan hóa Số lượng và Phân loại khách hàng 29](#_Toc23036)

[KẾT LUẬN 31](#_Toc23430)

[Kết quả đạt được. 31](#_Toc200)

[Hạn chế và hướng phát triển . 32](#_Toc14994)

[TÀI LIỆU THAM KHẢO 33](#_Toc23285)

**DANH MỤC HÌNH**

[Hình 1 : Bảng dữ liệu 9](#_Toc22440)

[Hình 2 : Bảng thông tin dữ liệu 10](#_Toc24668)

[Hình 3 : Các thư viện đã được sử dụng 11](#_Toc15652)

[Hình 4 : Nhóm các chỉ số theo năm 13](#_Toc26292)

[Hình 5 : Tổng Doanh thu theo năm(Revenue) 13](#_Toc20)

[Hình 6 : Tổng Lợi nhuận trước thuế theo năm(Profit\_Before\_Tax) 14](#_Toc28389)

[Hình 7 : Lợi nhuận ròng theo năm(Net\_Profit) 15](#_Toc25287)

[Hình 8 : Biên lợi nhuận theo năm(Profit\_Margin) 16](#_Toc11492)

[Hình 9 : Tổng nợ cho khách hàng vay 17](#_Toc4339)

[Hình 10 : Tổng nợ vay từ các tổ chức khác theo năm 18](#_Toc17196)

[Hình 11 : Số lượng và phân loại khách hàng 19](#_Toc8965)

[Hình 12 : Dữ liệu chuẩn bị cho phần học máy 20](#_Toc177)

[Hình 13 : Khởi tạo mô hình Random Forest 20](#_Toc27574)

[Hình 14 : Khởi tạo mô hình K-Nearest Neighbors 21](#_Toc15126)

[Hình 15 : Khởi tạo mô hình Decision Tree 21](#_Toc17375)

[Hình 16 : Kết quả của 3 mô hình 22](#_Toc6431)

[Hình 17 : Doanh thu dự đoán trong năm 2023 22](#_Toc18574)

[Hình 18 : Lợi nhuận trước thuế dự đoán theo năm 2023 23](#_Toc24937)

[Hình 19 : Lợi nhuận ròng dự đoán theo năm 2023 23](#_Toc16217)

[Hình 20 : Biên lợi nhuận dự đoán năm 2023 24](#_Toc8537)

[Hình 21 : Nợ cho khách hàng vay dự đoán năm 2023 24](#_Toc5362)

[Hình 22 : Nợ vay từ các tổ chức khác dự đoán năm 2023 24](#_Toc16466)

[Hình 23 : Trực quan hoá doanh thu qua các năm 25](#_Toc10748)

[Hình 24 : Trực quan hóa Lợi nhuận trước thuế theo năm 26](#_Toc21291)

[Hình 26 : Trực quan hóa Lợi nhuận ròng theo năm 27](#_Toc25117)

[Hình 26 : Trực quan hóa Biên lợi nhuận theo năm 28](#_Toc17376)

[Hình 28 : Trực quan hóa Các khoản vay của ngân hàng 29](#_Toc10581)

[Hình 29 : Trực quan hóa Số lượng và phân loại khách hàng 30](#_Toc8583)

**KÍ HIỆU CÁC CỤM TỪ VIẾT TẮT**

|  |  |
| --- | --- |
| Chữ viết tắt | Ý nghĩa |
| KNN | K-nearest neighbors |
| TMCP | Thương Mại Cổ Phần |
| VCB | Vietcombank |

# CHƯƠNG 1: GIỚI THIỆU

**1.1 Giới thiệu đề tài.**

## Trong bối cảnh chuyển đổi số ngày càng phát triển, ngành tài chính ngân hàng đang đối diện với sự bùng nổ dữ liệu, mang lại nhiều tiềm năng lớn trong việc tối ưu hóa quy trình quản lý và hỗ trợ ra quyết định. Đặc biệt đối với các ngân hàng lớn như Ngân hàng TMCP Ngoại Thương Việt Nam (VCB), việc phân tích và khai thác dữ liệu tài chính không chỉ giúp nâng cao khả năng kiểm soát hoạt động mà còn đưa ra những dự đoán giá trị về xu hướng thị trường và quản lý rủi ro trong tương lai. Tương tự như việc áp dụng Big Data trong lĩnh vực y tế để phát hiện sớm và quản lý các bệnh mãn tính như tiểu đường, công nghệ phân tích dữ liệu lớn trong ngành tài chính đang mang đến những thay đổi tích cực trong việc xử lý và phân tích khối lượng lớn dữ liệu tài chính.

## Đề tài "Phân tích dữ liệu về tài chính ngân hàng VCB" không chỉ tập trung vào việc sử dụng các phương pháp phân tích truyền thống, mà còn hướng đến việc cải thiện và nâng cao hiệu quả thông qua việc kết hợp với các công nghệ tiên tiến như Machine Learning. Điều này mở ra tiềm năng phát hiện các mẫu hình tiềm ẩn trong dữ liệu mà trước đây không thể nhận thấy, đồng thời hỗ trợ ngân hàng trong việc dự đoán chính xác xu hướng kinh doanh, tối ưu hóa việc quản lý rủi ro và nâng cao hiệu quả hoạt động.

## Với tiềm năng ứng dụng thực tiễn cao, đề tài không chỉ mang lại giá trị lớn cho nghiên cứu khoa học mà còn đóng góp trực tiếp vào việc nâng cao khả năng quản lý tài chính, ra quyết định chiến lược tại VCB. Điều này không chỉ giúp ngân hàng hoạt động hiệu quả hơn mà còn góp phần vào sự phát triển bền vững của hệ thống tài chính tại Việt Nam, đáp ứng tốt hơn nhu cầu của thị trường trong thời đại công nghệ phát triển mạnh mẽ.

## 1.2.Lý do chọn đề tài

Nếu phải nói lý do nào để khiến bọn em chọn đề tài "Phân tích dữ liệu về tài chính ngân hàng VCB" thì cự thể cũng không có nhiều lý do nhưng có thể kể đến những lý do như sau đó là:

* **Tính cấp bách trong phân tích dữ liệu tài chính:**

Trong bối cảnh nền kinh tế toàn cầu phát triển nhanh chóng và phức tạp, việc phân tích dữ liệu tài chính đã trở thành một công cụ thiết yếu để các ngân hàng tối ưu hóa hoạt động và hỗ trợ ra quyết định. Đối với Ngân hàng TMCP Ngoại Thương Việt Nam (VCB), việc phân tích dữ liệu đóng vai trò quan trọng không chỉ giúp hiểu rõ tình hình tài chính hiện tại mà còn dự đoán được xu hướng và rủi ro tiềm ẩn trong tương lai. Từ đó, ngân hàng có thể điều chỉnh chiến lược nhằm đảm bảo sự phát triển bền vững và nâng cao hiệu quả hoạt động.

* **Tiềm năng ứng dụng của Machine Learning trong phân tích tài chính:**

Machine Learning đã trở thành một công cụ mạnh mẽ trong việc phân tích và xử lý các tập dữ liệu lớn và phức tạp. Các thuật toán học máy có khả năng nhận diện các mẫu ẩn trong dữ liệu, từ đó hỗ trợ việc dự đoán chính xác các biến động trong lĩnh vực tài chính. Việc ứng dụng Machine Learning vào phân tích dữ liệu tại VCB có thể giúp ngân hàng đưa ra các quyết định chính xác hơn, từ quản lý rủi ro tín dụng đến tối ưu hóa danh mục đầu tư và dự báo xu hướng thị trường.

* **Tính ứng dụng cao và thực tiễn:**

Đề tài không chỉ dừng lại ở việc nghiên cứu lý thuyết mà còn mang tính ứng dụng cao trong thực tế. Các mô hình phân tích dữ liệu và Machine Learning có thể được áp dụng để cải thiện các quy trình tài chính tại VCB, từ đó giúp tối ưu hóa hoạt động, nâng cao năng suất và giảm thiểu rủi ro. Những kết quả của nghiên cứu này có tiềm năng đóng góp trực tiếp vào việc cải tiến hệ thống quản lý và hỗ trợ ra quyết định của ngân hàng.

* **Đóng góp cho cộng đồng khoa học và ngành tài chính:**

Nghiên cứu này không chỉ có giá trị thực tiễn trong việc phân tích dữ liệu tài chính mà còn đóng góp vào sự phát triển của lĩnh vực khoa học dữ liệu và công nghệ tài chính. Sử dụng Machine Learning trong phân tích dữ liệu mở ra những hướng đi mới trong việc quản lý tài chính, góp phần thúc đẩy các nghiên cứu chuyên sâu về ứng dụng công nghệ vào tài chính. Điều này không chỉ mang lại lợi ích cho ngân hàng mà còn có giá trị đối với cộng đồng nghiên cứu khoa học.

* **Hướng đến phát triển bền vững trong ngành tài chính:**

Phân tích chính xác và dự đoán hiệu quả các xu hướng tài chính giúp các ngân hàng như VCB xây dựng chiến lược phát triển bền vững. Điều này không chỉ đảm bảo giảm thiểu rủi ro mà còn giúp tối ưu hóa nguồn lực, từ đó đóng góp vào sự phát triển dài hạn và ổn định của ngân hàng. Nghiên cứu này sẽ góp phần quan trọng trong việc xây dựng một nền tảng tài chính vững chắc, giúp ngân hàng VCB tiếp tục phát triển mạnh mẽ trong môi trường kinh doanh hiện đại.

## 1.3.Mục tiêu đề tài.

Mục tiêu mà em đề ra cho đề tài "Phân tích dữ liệu về tài chính ngân hàng VCB" bao gôm các mục tiêu sau:

* **Xây dựng mô hình dự báo chính xác:** Tạo ra các mô hình phân tích dựa trên dữ liệu tài chính nhằm giúp ngân hàng VCB đánh giá các yếu tố rủi ro và dự đoán xu hướng thị trường trong tương lai. Mục tiêu là xây dựng các công cụ dự đoán với độ chính xác cao, hỗ trợ quá trình ra quyết định tài chính hiệu quả.
* **Phân tích các yếu tố rủi ro chủ yếu:** Nghiên cứu kỹ lưỡng và đánh giá ảnh hưởng của các yếu tố rủi ro quan trọng như rủi ro tín dụng, thị trường và thanh khoản đối với hoạt động ngân hàng. Qua đó, VCB có thể đề xuất các giải pháp quản lý rủi ro, nâng cao năng lực kiểm soát và giảm thiểu rủi ro tài chính.
* **Cung cấp cơ sở cho các quyết định chiến lược:** Đề tài hướng đến việc cung cấp những thông tin phân tích sâu sắc và các dự đoán tài chính đáng tin cậy nhằm hỗ trợ các quyết định chiến lược như quản lý tài sản, tối ưu hóa danh mục đầu tư và tìm kiếm cơ hội phát triển. Điều này sẽ giúp tăng cường tính hiệu quả trong việc hoạch định chiến lược của ngân hàng.
* **Phòng ngừa và quản lý rủi ro tài chính:** Sử dụng các phân tích dữ liệu để đưa ra các biện pháp phòng ngừa và quản lý rủi ro, nhằm giúp VCB chủ động ứng phó với các biến động trên thị trường tài chính. Mục tiêu là giảm thiểu tác động tiêu cực lên hoạt động của ngân hàng, từ đó cải thiện hiệu suất kinh doanh.
* **Ứng dụng trên nhiều phân khúc khách hàng:** Thực hiện phân tích dữ liệu tài chính trên nhiều nhóm khách hàng khác nhau, bao gồm khách hàng cá nhân, doanh nghiệp và các tổ chức đối tác. Điều này giúp ngân hàng nắm bắt được nhu cầu và hành vi của các nhóm đối tượng, từ đó điều chỉnh các chiến lược kinh doanh cho phù hợp.

Tóm lại, nghiên cứu này đặt mục tiêu xây dựng các công cụ phân tích dữ liệu tài chính hữu ích, hỗ trợ ngân hàng VCB ra quyết định, nâng cao hiệu quả quản lý rủi ro và phát triển bền vững trong môi trường kinh doanh đầy thách thức.

## 1.4.Phương pháp đề tài.

Có một số phương pháp tiếp cận đề tài "Phân tích dữ liệu về tài chính ngân hàng VCB" mà em sử dụng để đạt được mục tiêu của nó

**Thu thập và xử lý dữ liệu:** Dữ liệu được lấy từ nhiều nguồn khác nhau, bao gồm báo cáo tài chính, thông tin giao dịch ngân hàng và các yếu tố kinh tế vĩ mô. Quá trình xử lý dữ liệu sẽ bao gồm việc làm sạch, xử lý dữ liệu bị thiếu và chuẩn hóa nhằm đảm bảo dữ liệu có chất lượng cao, đồng thời sẵn sàng để phân tích. Đây là bước quan trọng để đảm bảo tính chính xác và đáng tin cậy của các mô hình phân tích.

**Phân tích dữ liệu và xây dựng mô hình:** Sử dụng các kỹ thuật phân tích dữ liệu như hồi quy, cây quyết định, rừng ngẫu nhiên và thuật toán Machine Learning nhằm khai thác thông tin từ dữ liệu tài chính. Các mô hình này sẽ giúp dự đoán các yếu tố như rủi ro tín dụng, biến động thị trường và tình hình tài chính của ngân hàng. Phân tích Big Data cũng được áp dụng để xử lý khối lượng dữ liệu lớn và phức tạp một cách hiệu quả.

**Đánh giá và cải thiện mô hình:** Áp dụng các phương pháp như cross-validation, confusion matrix và phân tích đường cong ROC để đánh giá hiệu suất của các mô hình. Dựa trên các chỉ số hiệu suất, các mô hình sẽ được tối ưu hóa bằng cách điều chỉnh các tham số và áp dụng các kỹ thuật chọn lọc đặc trưng nhằm cải thiện khả năng dự đoán và độ chính xác.

**Xác định các yếu tố rủi ro quan trọng:** Sử dụng các phương pháp phân tích như feature importance, SHAP values hoặc permutation importance để xác định các yếu tố có ảnh hưởng lớn đến hoạt động tài chính và rủi ro của ngân hàng. Việc nhận diện các yếu tố này giúp ngân hàng chủ động hơn trong việc quản lý và kiểm soát rủi ro.

**Ứng dụng và triển khai:** Sau khi hoàn thiện các mô hình phân tích, chúng sẽ được ứng dụng vào việc dự đoán xu hướng tài chính, quản lý rủi ro và hỗ trợ ra quyết định chiến lược cho VCB. Những kết quả từ mô hình sẽ giúp ngân hàng tối ưu hóa quy trình quản lý tài chính và đưa ra các quyết định kinh doanh phù hợp trong môi trường thị trường đầy biến động.

Bằng cách kết hợp các phương pháp này, đề tài sẽ cung cấp một quy trình phân tích toàn diện, giúp ngân hàng VCB cải thiện khả năng quản lý tài chính, giảm thiểu rủi ro và tăng cường hiệu quả hoạt động.

## 1.5.Đối tượng và phạm vi nghiên cứu.

Đối tượng và phạm vi nghiên cứu trong dự án về phân tích dữ liệu tài chính ngân hàng VCB có thể được xác định như sau:

### **1.5.1. Đối Tượng Nghiên Cứu**

Đối tượng nghiên cứu chính là các dữ liệu tài chính của Ngân hàng TMCP Ngoại Thương Việt Nam (VCB), bao gồm thông tin liên quan đến khách hàng, các giao dịch tài chính, tín dụng, và các chỉ số hoạt động tài chính khác của ngân hàng. Phân tích các yếu tố này sẽ giúp xây dựng các mô hình nhằm dự đoán rủi ro tài chính và hỗ trợ quá trình ra quyết định chiến lược cho VCB.

### **1.5.2. Phạm Vi Nghiên Cứu**

**Thu thập và sử dụng dữ liệu**: Dữ liệu được thu thập từ nhiều nguồn, bao gồm hệ thống nội bộ của VCB, các báo cáo tài chính và thông tin giao dịch. Sau khi thu thập, dữ liệu sẽ được làm sạch, loại bỏ các giá trị thiếu và chuẩn hóa để đảm bảo tính đồng nhất, sẵn sàng cho các bước phân tích tiếp theo.

**Xây dựng mô hình dự đoán**: Nghiên cứu tập trung vào phát triển các mô hình phân tích và dự đoán dựa trên các thuật toán học máy như hồi quy, rừng ngẫu nhiên và cây quyết định. Các mô hình này sẽ giúp ngân hàng dự đoán các yếu tố quan trọng như rủi ro tín dụng, hiệu suất tài chính, và xu hướng thị trường.

**Đánh giá và tinh chỉnh mô hình**: Hiệu suất của mô hình sẽ được đánh giá và các siêu tham số sẽ được tinh chỉnh để tối ưu hóa khả năng dự đoán.

**Ứng dụng trong y học thực tế**: Khi các mô hình đã được xây dựng và tối ưu hóa, chúng sẽ được triển khai vào quá trình hoạt động của ngân hàng. Các mô hình này sẽ hỗ trợ VCB trong việc quản lý rủi ro, đưa ra quyết định chiến lược và cải thiện hiệu suất tài chính, giúp ngân hàng hoạt động hiệu quả hơn trong môi trường kinh doanh biến động.

Phạm vi nghiên cứu bao gồm toàn bộ quy trình từ thu thập, xử lý dữ liệu đến xây dựng và triển khai các mô hình dự đoán. Mục tiêu là cung cấp những công cụ hữu ích cho VCB trong việc quản lý tài chính và ra quyết định chiến lược.

# CHƯƠNG 2 ỨNG DỤNG VÀ THUẬT TOÁN

## 2.1 Giới thiệu bài toán.

Trong bối cảnh thị trường tài chính ngày càng biến động, việc quản lý rủi ro và tối ưu hóa hoạt động ngân hàng đang trở thành một thách thức quan trọng đối với các tổ chức tài chính, bao gồm Ngân hàng TMCP Ngoại Thương Việt Nam (VCB). Số lượng giao dịch tăng cao cùng với những biến động khó lường của thị trường đòi hỏi các ngân hàng phải đưa ra quyết định nhanh chóng và chính xác. Điều này khiến việc dự đoán các rủi ro tiềm ẩn và xu hướng thị trường trở thành một bài toán cấp thiết.

Bài toán chính ở đây là phát triển các mô hình dự báo có khả năng phân tích và dự đoán sớm những yếu tố rủi ro quan trọng, bao gồm rủi ro tín dụng, biến động thị trường và hiệu quả hoạt động tài chính. Mục tiêu của nghiên cứu là sử dụng dữ liệu từ các giao dịch, báo cáo tài chính, và các chỉ số thị trường để đưa ra dự đoán đáng tin cậy, hỗ trợ quá trình ra quyết định của ngân hàng.

Trong quá trình này, công nghệ Machine Learning và phân tích dữ liệu lớn (Big Data) trở thành những công cụ then chốt. Nhờ vào khả năng xử lý khối lượng lớn dữ liệu và phát hiện ra những mẫu ẩn sâu trong dữ liệu, các thuật toán học máy có thể giúp ngân hàng phân tích hiệu quả hơn các yếu tố rủi ro, điều mà các phương pháp phân tích truyền thống không thể làm được. Điều này giúp nâng cao khả năng dự đoán và tối ưu hóa quá trình quản lý tài chính của VCB.

Mục tiêu chính của bài toán là xây dựng các mô hình dự báo rủi ro tài chính, phân tích hiệu suất kinh doanh và đưa ra các dự đoán xu hướng thị trường cho ngân hàng VCB. Việc dự đoán sớm những yếu tố rủi ro tiềm ẩn sẽ giúp ngân hàng có thể chủ động triển khai các biện pháp phòng ngừa, từ đó giảm thiểu tổn thất và nâng cao hiệu quả hoạt động kinh doanh.

Ngoài giá trị về mặt nghiên cứu, bài toán còn có tiềm năng ứng dụng cao trong thực tế. Các công cụ và mô hình dự đoán sẽ hỗ trợ ngân hàng cải thiện quy trình ra quyết định và nâng cao hiệu suất quản lý rủi ro, đồng thời góp phần vào sự phát triển bền vững của VCB trong môi trường tài chính đầy biến động.

## 2.2 Mô tả thuật toán.

## **2.2.1. Thuật toán Disision Tree**

**Thuật toán Decision Tree (Cây quyết định) là một công cụ mạnh mẽ trong phân tích dữ liệu tài chính, đặc biệt trong việc xử lý các tình huống như đánh giá rủi ro tín dụng hoặc phân tích hiệu suất tài chính tại Ngân hàng TMCP Ngoại Thương Việt Nam (VCB). Decision Tree hoạt động bằng cách phân loại dữ liệu thành các nhóm nhỏ dựa trên các yếu tố quan trọng, như tình hình tín dụng của khách hàng, xu hướng thị trường, hoặc biến động lãi suất. Tại mỗi giai đoạn, thuật toán sẽ lựa chọn thuộc tính tài chính tối ưu nhất để tiếp tục phân tách dữ liệu, giúp hình thành một chuỗi các quyết định logic rõ ràng.**

**Thuật toán này không chỉ giúp dự báo các rủi ro tài chính mà còn hỗ trợ đưa ra các quyết định quan trọng liên quan đến quản lý tín dụng và lập chiến lược đầu tư. Với tính chất dễ hiểu và trực quan, Decision Tree cung cấp cho các nhà quản lý tại VCB một công cụ mạnh mẽ để theo dõi và phân tích dữ liệu, giúp họ dễ dàng hiểu rõ quy trình ra quyết định và thực hiện các hành động chiến lược phù hợp.**

### **2.2.2. Thuật toán Random Forest**

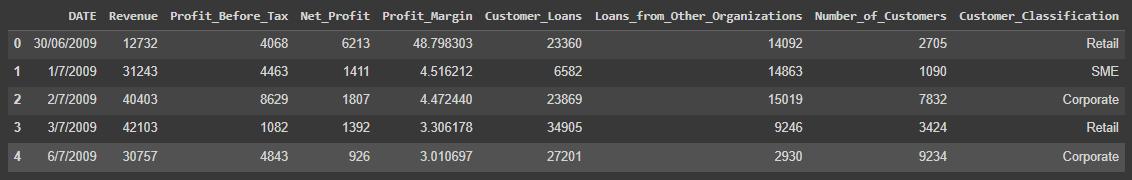
**Random Forest** là một thuật toán học máy kết hợp nhiều cây quyết định để cải thiện độ chính xác trong các dự đoán tài chính. Thuật toán này được thiết kế để xử lý tốt các tập dữ liệu lớn và phức tạp, giúp giảm thiểu hiện tượng quá khớp (overfitting) khi mô hình học quá mức từ dữ liệu. Trong phân tích dữ liệu tài chính tại ngân hàng, Random Forest giúp nhận diện các yếu tố quan trọng nhất ảnh hưởng đến kết quả dự báo, chẳng hạn như rủi ro tín dụng, biến động lãi suất, hoặc xu hướng thị trường. Kỹ thuật đánh giá ngoài túi (Out-of-Bag Evaluation) và quá trình điều chỉnh các siêu tham số là những công cụ hữu ích để tối ưu hóa hiệu suất của mô hình, đảm bảo mô hình hoạt động hiệu quả và chính xác, hỗ trợ tốt hơn cho việc ra quyết định chiến lược của ngân hàng.

### **2.2.3. Thuật toán K-Nearest Neighbors**

Thuật toán K-Nearest Neighbors (KNN) phân loại các dữ liệu tài chính dựa trên những điểm dữ liệu gần nhất trong không gian đặc trưng. Trong bối cảnh dự đoán rủi ro tài chính tại ngân hàng VCB, KNN xác định các điểm dữ liệu tương tự nhất và từ đó đưa ra dự đoán liên quan đến các yếu tố như rủi ro tín dụng hoặc hiệu suất tài chính. Mặc dù KNN không trực tiếp cung cấp xác suất, có thể ước tính được xác suất thông qua tỷ lệ của các điểm láng giềng thuộc các nhóm khác nhau. Việc lựa chọn số lượng láng giềng (K) đóng vai trò quan trọng trong hiệu suất của mô hình, cần điều chỉnh một cách hợp lý để tránh tình trạng mô hình quá khớp hoặc không đủ khớp, nhằm đảm bảo kết quả chính xác và hữu ích cho việc ra quyết định.

## 2.3. Xây dựng bộ dữ liệu.

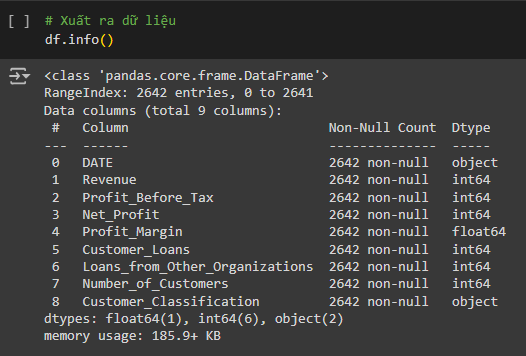
### **2.3.1. Bộ dữ liệu gồm các cột**



Hình 1: Bảng dữ liệu

1. **DATE:** Ngày báo cáo tài chính
2. **Revenue**: Tổng doanh thu
3. **Profit\_Before\_Tax**: Lợi nhuận trước thuế của công ty
4. **Net\_Profit**: Lợi nhuận ròng (lợi nhuận sau thuế)
5. **Profit\_Margin**: Biên lợi nhuận
6. **Customer\_Loans**: Tổng các khoản vay của khách hàng
7. **Loans\_from\_Other\_Organizations**: Các khoản vay từ các tổ chức khác
8. **Number\_of\_Customers**: Số lượng khách hàng hiện tại của công ty
9. **Customer\_Classification**: Phân loại khách hàng theo các nhóm

### **2.3.2. Tiền xử lý dữ liệu**



Hình 2: Bảng thông tin dữ liệu

## 2.4. Áp dụng thuật toán vào bài toán.

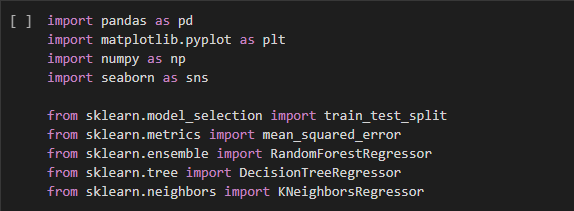
Trong quá trình nghiên cứu đề tài "Phân tích dữ liệu tài chính ngân hàng VCB", tôi đã sử dụng ba thuật toán học máy quan trọng để thực hiện các dự báo và phân tích dữ liệu. Mỗi thuật toán mang lại những lợi ích khác nhau trong việc xử lý và phân tích dữ liệu tài chính.

* **Random Forest**: Đây là một phương pháp kết hợp nhiều cây quyết định lại với nhau nhằm tăng độ chính xác trong dự báo. Random Forest nổi bật nhờ khả năng xử lý dữ liệu phức tạp và giảm thiểu vấn đề overfitting. Với dữ liệu tài chính ngân hàng, thuật toán này rất hữu ích trong việc phân loại và dự báo kết quả dựa trên các yếu tố đa dạng và phức tạp.
* **Decision Tree**: Cây quyết định là một công cụ mạnh mẽ, giúp phân tích dữ liệu và đưa ra dự báo dựa trên các đặc trưng cụ thể. Nhờ tính dễ hiểu và trực quan, thuật toán này cho phép xây dựng các quy tắc dựa trên các yếu tố ảnh hưởng trực tiếp đến kết quả kinh doanh, giúp người dùng dễ dàng theo dõi quy trình phân tích và đưa ra quyết định dựa trên dữ liệu tài chính.
* **K-nearest neighbors (KNN)**: KNN hoạt động bằng cách tính toán khoảng cách giữa các điểm dữ liệu trong không gian đặc trưng và dự đoán dựa trên những điểm tương tự nhất. Khi ứng dụng vào lĩnh vực tài chính ngân hàng, KNN hỗ trợ việc phân loại khách hàng và phát hiện các mẫu bất thường thông qua so sánh dữ liệu hiện tại với những mẫu đã có.

Những thuật toán này không chỉ giúp tôi hiểu rõ hơn về các mô hình dữ liệu tài chính mà còn mang lại khả năng dự báo chính xác, từ đó hỗ trợ tối ưu hóa các quyết định trong hoạt động tài chính của ngân hàng VCB.

## 2.5. Thực nghiệm với thư viện.

Hầu hết những ai đã từng học lập trình đều không còn xa lạ với thuật ngữ "Modules" – ám chỉ các thư viện lập trình chứa những đoạn mã tái sử dụng, được phát triển bởi các lập trình viên. Để chuẩn bị cho đề tài "Phân tích dữ liệu về tài chính ngân hàng VCB," việc sử dụng các thư viện lập trình là một phần không thể thiếu. Dưới đây là một số thư viện đã được tôi áp dụng trong quá trình thực hiện dự án. Các thư viện này không chỉ giúp rút ngắn thời gian phát triển mà còn cung cấp nhiều công cụ hữu ích cho việc phân tích dữ liệu và xây dựng các mô hình dự đoán.



Hình 3: Các thư viện đã được sử dụng

**1.pandas (pd):**

Chúng ta bắt đầu với một thư viện rất quen thuộc với những người mới học Python. Đúng rồi, thư viện mà mình muốn nhắc đến chính là Pandas. Đây là một công cụ mạnh mẽ để xử lý và phân tích dữ liệu, cung cấp các cấu trúc dữ liệu như DataFrame và Series, giúp việc thao tác dữ liệu và trực quan hóa trở nên dễ dàng hơn.

**2.matplotlib.pyplot (plt):**

Thư viện tiếp theo mà em muốn giới thiệu là matplotlib, một công cụ hữu ích trong việc vẽ đồ thị trên Python.

Pyplot là một module của matplotlib, cung cấp các hàm tạo và hiển thị biểu đồ với các giao diện như Tkinter và wxPython.

**3.seaborn (sns):**

Seaborn là một thư viện trực quan hóa dữ liệu được xây dựng dựa trên matplotlib. Thư viện này cung cấp giao diện thân thiện, giúp dễ dàng tạo ra các biểu đồ thống kê với tính trực quan cao.

**4.numpy (np):**

Numpy là một thư viện quan trọng trong Python, chuyên dùng cho các phép tính số học. Nó cung cấp các đối tượng mảng và nhiều hàm toán học mạnh mẽ, giúp quá trình xử lý và thao tác dữ liệu trở nên nhanh chóng và hiệu quả hơn.

**5.scikit-learn**

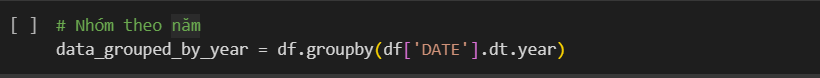
Scikit-learn, hay còn gọi tắt là sklearn, là một trong những thư viện phổ biến nhất trong lĩnh vực machine learning và thống kê. Nó cung cấp các công cụ mạnh mẽ để xây dựng, đánh giá mô hình học máy, và được sử dụng rộng rãi trên các hệ điều hành như Linux nhờ vào giấy phép bản quyền FreeBSD.

**6. plotly**

Plotly là một thư viện đồ họa tương tác, mã nguồn mở, giúp tạo ra các biểu đồ tương tác trên nền tảng trình duyệt. Thư viện này rất phù hợp cho việc trực quan hóa dữ liệu với khả năng tương tác cao.

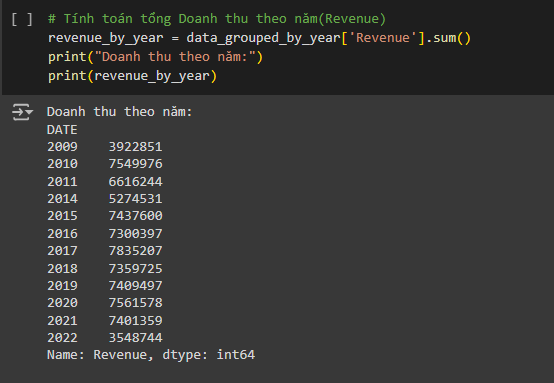
# CHƯƠNG 3 XÂY DỰNG ỨNG DỤNG NGÔN NGỮ PYTHON

## 3.1 Xem xét các chỉ số lợi nhuận



Hình 4: Nhóm các chỉ số theo năm

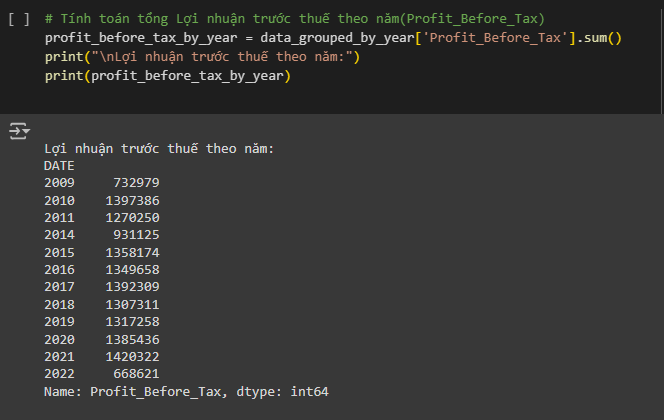
**3.1.1. Doanh Thu**



Hình 5: Tổng Doanh thu theo năm(Revenue)

Doanh thu của ngân hàng từ năm 2009 đến 2022 thể hiện sự dao động đáng kể. Năm 2009 ghi nhận mức doanh thu thấp nhất (****3,922,851**** đơn vị), nhưng trong giai đoạn 2010-2013, doanh thu tăng trưởng mạnh, đặc biệt là năm 2010 với ****7,549,976**** đơn vị. Năm 2014 chứng kiến sự sụt giảm doanh thu xuống còn ****5,274,531**** đơn vị. Tuy nhiên, từ năm 2015 đến 2021, doanh thu dần phục hồi và đạt mức cao nhất vào năm 2021 (****7,401,359**** đơn vị). Đến năm 2022, doanh thu giảm mạnh xuống ****3,548,744**** đơn vị. Sự biến động này có thể do ảnh hưởng từ các yếu tố kinh tế hoặc chiến lược của ngân hàng.

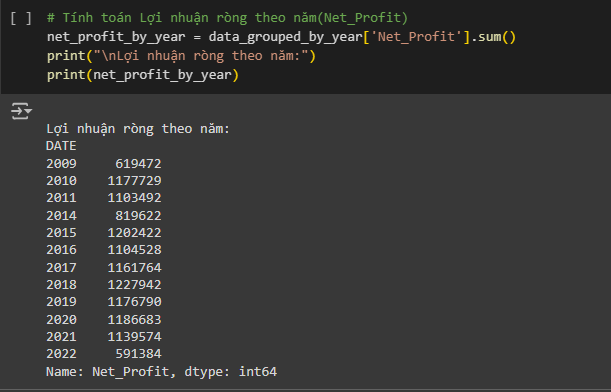
### **3.1.2. Lợi nhuận trước thuế**



Hình 6: Tổng Lợi nhuận trước thuế theo năm(Profit\_Before\_Tax)

Lợi nhuận trước thuế của ngân hàng từ năm 2009 đến 2022 có sự dao động đáng chú ý. Năm 2009, lợi nhuận đạt **732,979** đơn vị và liên tục tăng qua các năm, đạt mức cao nhất vào năm 2018 với **1,393,290** đơn vị. Tuy nhiên, từ năm 2019 trở đi, lợi nhuận có xu hướng giảm, đặc biệt năm 2022 chỉ đạt **668,621** đơn vị. Sự sụt giảm này có thể là kết quả của sự thay đổi trong chiến lược kinh doanh hoặc tác động từ các yếu tố kinh tế bên ngoài.

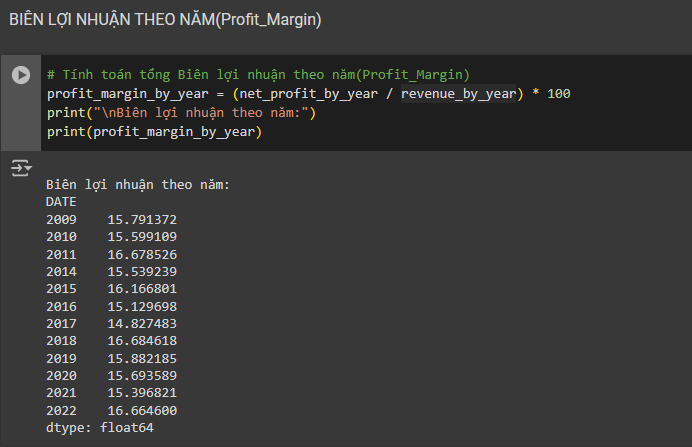
### **3.1.3. Lợi nhuận ròng**



Hình 7: Lợi nhuận ròng theo năm(Net\_Profit)

Lợi nhuận ròng của ngân hàng từ năm 2009 đến 2022 thể hiện sự biến động rõ rệt. Năm 2009, ngân hàng ghi nhận **6,194,72** đơn vị lợi nhuận ròng, và con số này tăng dần trong những năm sau, đạt mức cao nhất vào năm 2019 với **11,767,90** đơn vị. Tuy nhiên, từ năm 2020 trở đi, lợi nhuận bắt đầu suy giảm, xuống còn **5,913,84** đơn vị vào năm 2022. Xu hướng sụt giảm này có thể là do tác động từ các yếu tố kinh tế hoặc điều chỉnh chiến lược kinh doanh của ngân hàng.

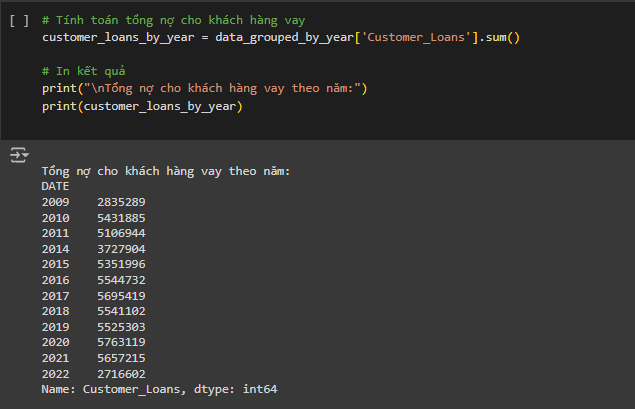
**3.1.4. Biên lợi nhuận**



Hình 8: Biên lợi nhuận theo năm(Profit\_Margin)

Biên lợi nhuận của ngân hàng từ năm 2009 đến 2022 có sự thay đổi nhất định nhưng vẫn nằm trong khoảng ổn định. Năm 2009, biên lợi nhuận đạt **15.79%**, sau đó tăng lên mức cao nhất là **15.88%** vào năm 2019. Từ năm 2020, biên lợi nhuận bắt đầu có dấu hiệu giảm, xuống mức **12.66%** vào năm 2022. Sự sụt giảm này có thể là kết quả của những biến động kinh tế hoặc điều chỉnh trong chiến lược hoạt động của ngân hàng.

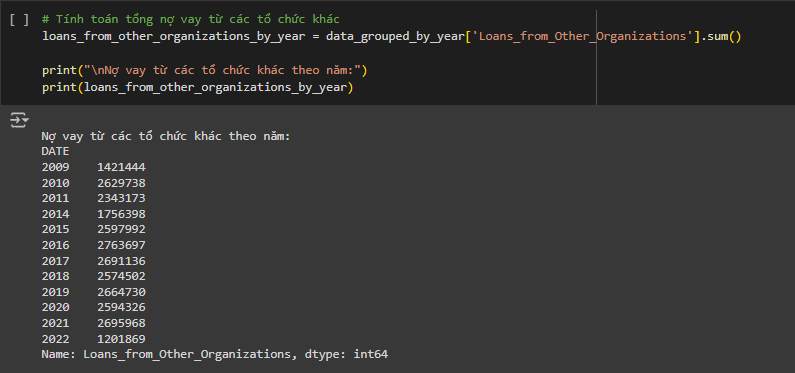
**3.1.5. Tổng nợ cho khách hàng vay**



Hình 9: Tổng nợ cho khách hàng vay

Trong giai đoạn từ năm 2009 đến 2022, tổng nợ vay của khách hàng tại ngân hàng có xu hướng tăng dần, ngoại trừ một số năm giảm nhẹ. Năm 2009, tổng nợ cho vay đạt **2,835,289** đơn vị và tiếp tục tăng lên, đạt đỉnh vào năm 2021 với **5,657,215** đơn vị. Tuy nhiên, đến năm 2022, con số này giảm xuống còn **2,716,602** đơn vị. Xu hướng tăng trưởng qua các năm cho thấy sự mở rộng tín dụng của ngân hàng, trong khi sự giảm sút trong năm 2022 có thể phản ánh các thay đổi về chính sách hoặc tác động từ môi trường kinh tế vĩ mô.

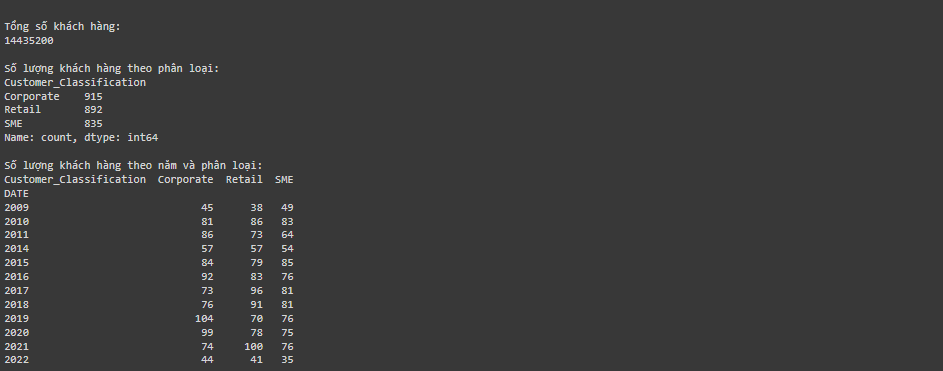
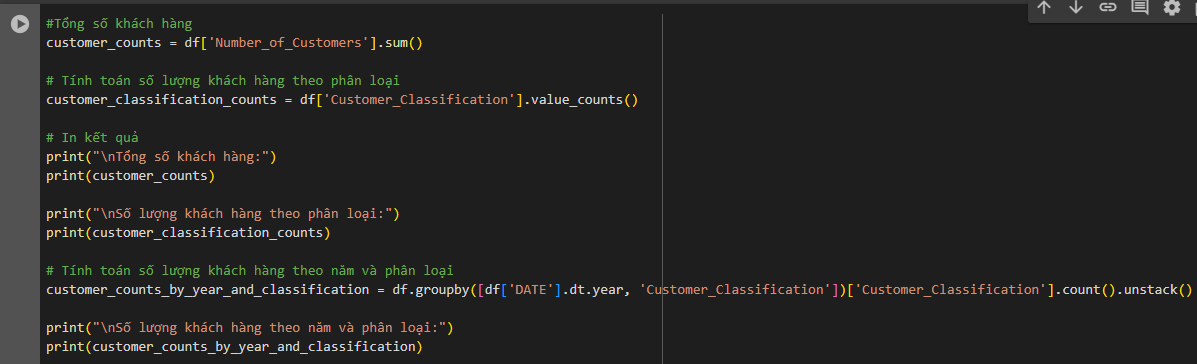
**3.1.6. Tổng nợ vay từ các tổ chức khác**



Hình 10: Tổng nợ vay từ các tổ chức khác theo năm

Nợ vay từ các tổ chức khác của ngân hàng trong giai đoạn từ 2009 đến 2022 thể hiện sự biến động. Bắt đầu từ **1,421,444** đơn vị vào năm 2009, tổng nợ vay tăng dần qua các năm, đạt đỉnh vào năm 2021 với **2,695,986** đơn vị. Tuy nhiên, đến năm 2022, con số này giảm xuống còn **1,201,869** đơn vị. Sự biến động này có thể là kết quả của những điều chỉnh trong chiến lược vay vốn của ngân hàng hoặc ảnh hưởng từ các điều kiện kinh tế bên ngoài.

**3.1.7. Số lượng và phân loại khách hàng**



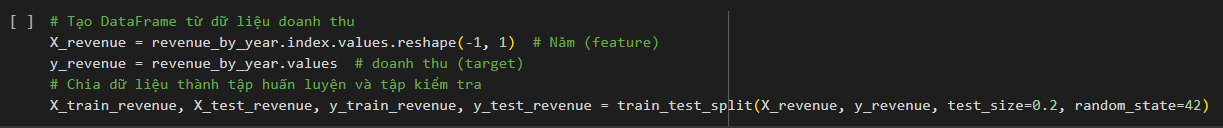
Hình 11: Số lượng và phân loại khách hàng

Tổng số khách hàng của ngân hàng từ năm 2009 đến 2022 là **145,228** khách hàng, được phân thành ba nhóm chính: **Corporate** (doanh nghiệp), **Retail** (bán lẻ) và **SME** (doanh nghiệp vừa và nhỏ). Trong đó, nhóm **Corporate** chiếm tỷ trọng lớn nhất với **85,902** khách hàng, tiếp theo là nhóm **Retail** với **58,521** khách hàng, và cuối cùng là nhóm **SME** với **805** khách hàng.

Số lượng khách hàng trong các phân loại cũng có sự thay đổi qua từng năm. Nhóm **Corporate** và **Retail** có xu hướng tương đối ổn định từ năm 2015 đến 2021, trong khi số lượng khách hàng của nhóm **SME** duy trì ở mức thấp và ít biến động. Năm 2022 ghi nhận sự giảm nhẹ trong cả ba nhóm khách hàng.

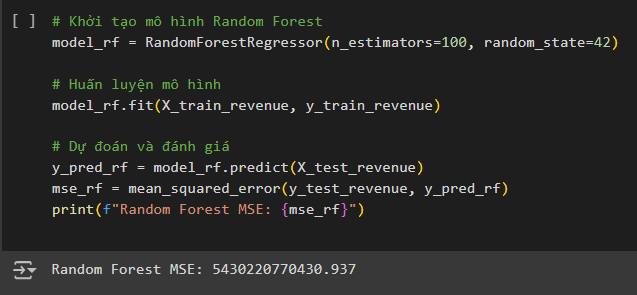
## 3.2 Xây dựng ứng dụng và giải thích

Trước khi đi vào các thuật toán dự đoán thì em sẽ dùng 3 mô hình là Random Forest, Dicision Tree và K Neighbors Regressor dự đoán doanh thu để chọn ra mô hình có hiệu suất tốt nhất :



Hình 12: Dữ liệu chuẩn bị cho phần học máy

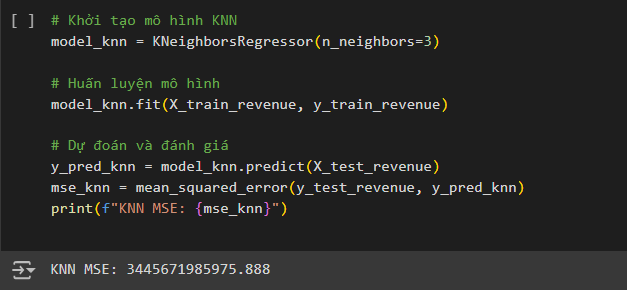
### **3.2.1. Random Forest**



Hình 13: Khởi tạo mô hình Random Forest

Như mọi người có thể thấy được ở phần kết quả của bài dự đoán của mô hình Random Forest chiếm 5430228770480.937

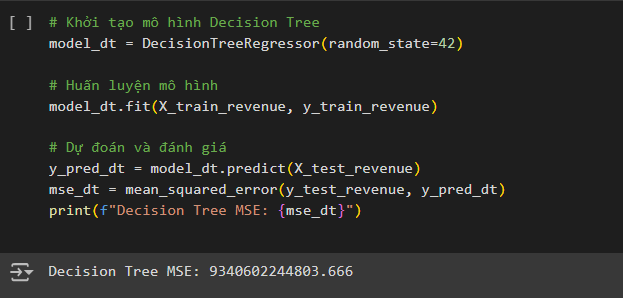
### **3.2.2. K-Nearest Neighbors**



Hình 14: Khởi tạo mô hình K-Nearest Neighbors

Như mọi người có thể thấy được ở phần kết quả của bài dự đoán của mô hình K-Nearest Neighbors chiếm 3445671985975.888

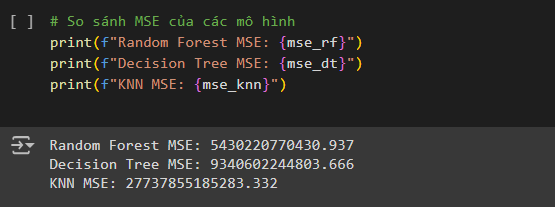
### **3.2.3. Decision Tree Classifier**



Hình 15: Khởi tạo mô hình Decision Tree

Như mọi người có thể thấy được ở phần kết quả của bài dự đoán của mô hình Decision Tree chiếm 9340602244803.666

**3.2.4. So sanh các phương pháp dự đoán**

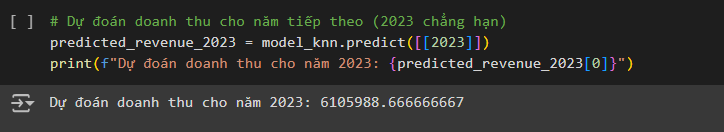


Hình 16: Kết quả của 3 mô hình

Từ 3 mô hình dự đoán bao gồm: K-Nearest Neighbors, Random Forest, Decision Tree Classifier. Ta có thể thấy được rằng mô hình K-Nearest Neighbors là mô hình có khả năng dự đoán tốt nhất nên ta sẽ sử dụng K-Nearest Neighbors đối với các đối tượng còn lại

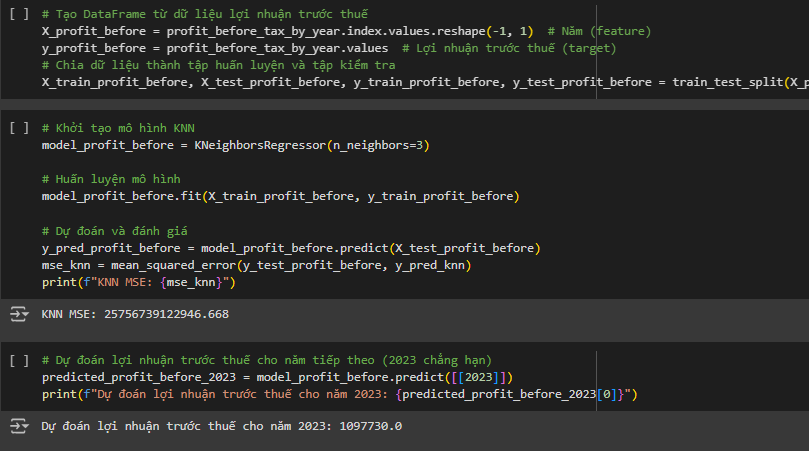
**3.3. Dự đoán các chỉ số cho tương lai**

**3.3.1. Dự đoán doanh thu**



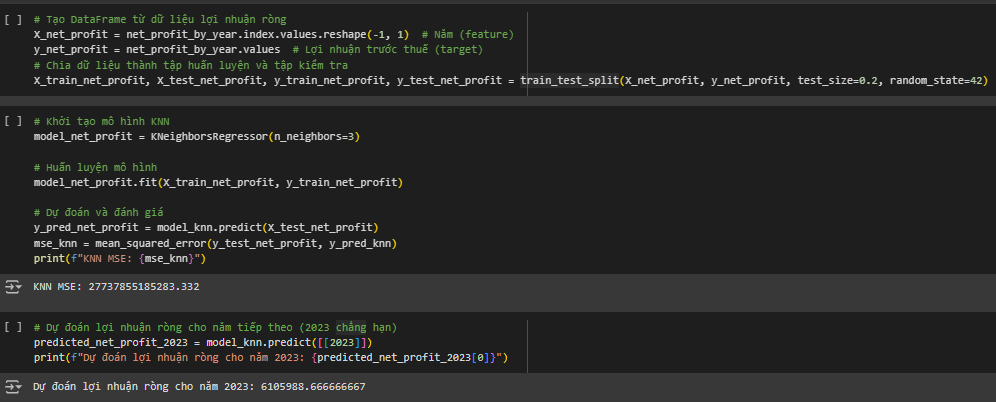
Hình 17: Doanh thu dự đoán trong năm 2023

**3.3.2. Dự đoán Lợi nhuận trước thuế**



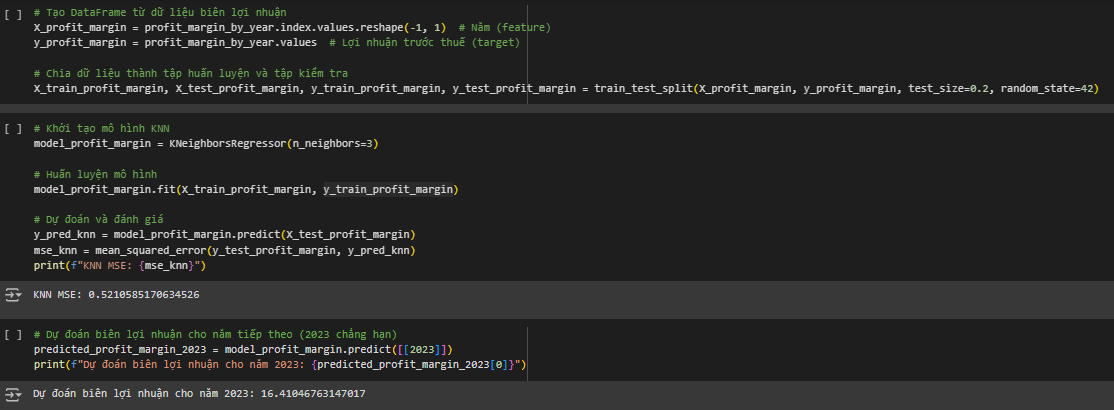
Hình 18: Lợi nhuận trước thuế dự đoán theo năm 2023

**3.3.3. Dự đoán lợi nhuận ròng**



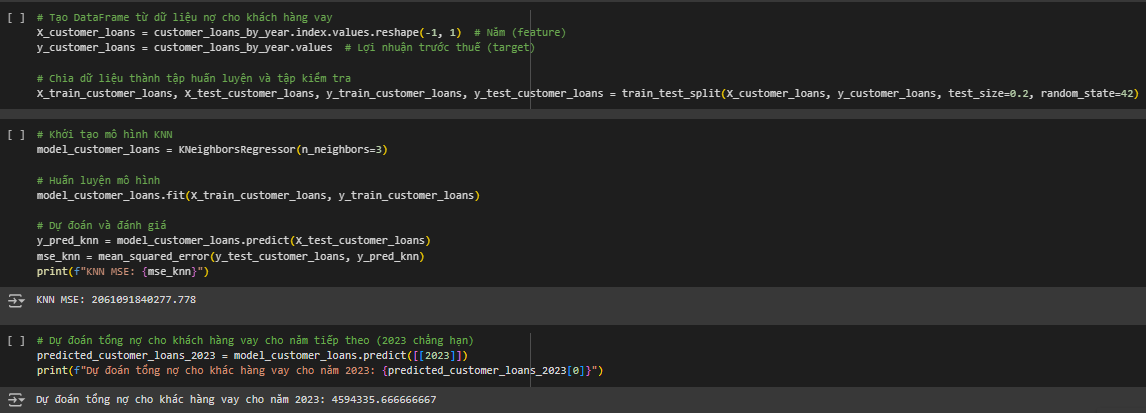
Hình 19: Lợi nhuận ròng dự đoán theo năm 2023

**3.3.4. Dự đoán biên lợi nhuận**



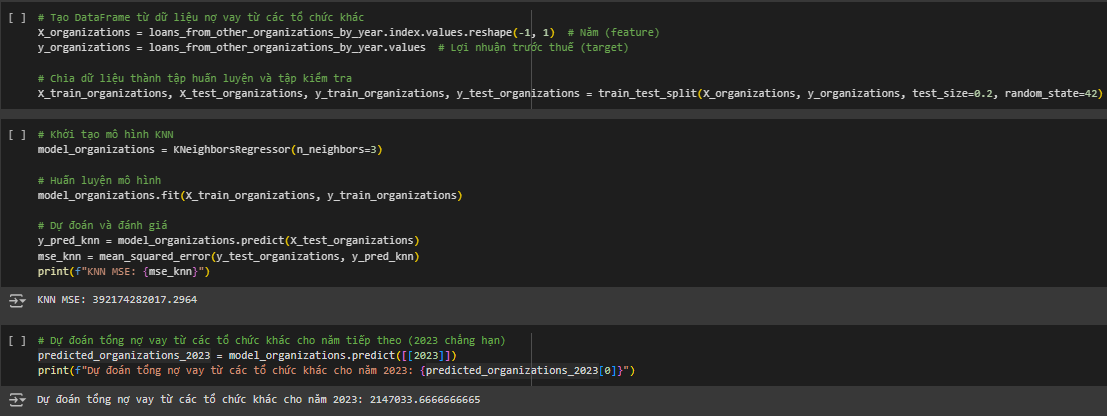
Hình 20: Biên lợi nhuận dự đoán năm 2023

**3.3.5. Dự đoán Nợ cho khách hàng vay**



Hình 21: Nợ cho khách hàng vay dự đoán năm 2023

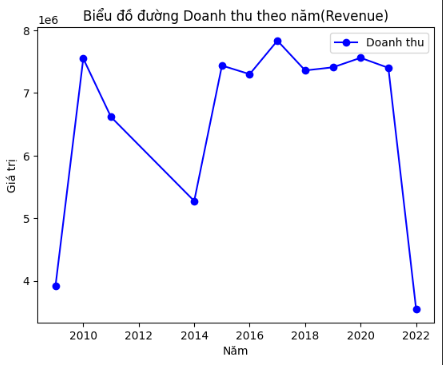
**3.3.6. Dự đoán Nợ vay từ các tổ chức khác**



Hình 22: Nợ vay từ các tổ chức khác dự đoán năm 2023

**3.4. Một số ảnh trực phân hóa được phân tích trong dự án**

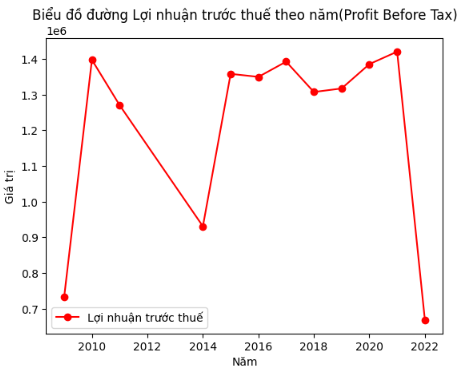
**3.4.1. Trực quan hoá doanh thu(Revenue)**



Hình 23: Trực quan hoá doanh thu qua các năm

Biểu đồ doanh thu từ năm 2009 đến 2022 cho thấy sự dao động đáng kể. Sau khi tăng mạnh và đạt đỉnh vào năm 2010, doanh thu giảm sút rõ rệt vào năm 2014. Từ năm 2015, doanh thu bắt đầu hồi phục và duy trì ở mức ổn định đến năm 2020. Tuy nhiên, đến năm 2022, doanh thu sụt giảm đáng kể, đánh dấu mức thấp nhất trong suốt giai đoạn theo dõi. Những thay đổi này có thể xuất phát từ sự điều chỉnh chiến lược của ngân hàng hoặc các yếu tố bên ngoài ảnh hưởng đến hoạt động kinh doanh.

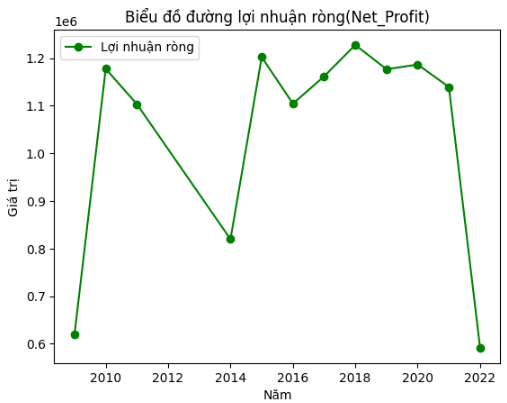
**3.4.2. Trực quan hóa Lợi nhuận trước thuế**



Hình 24: Trực quan hóa Lợi nhuận trước thuế theo năm

Biểu đồ lợi nhuận trước thuế từ năm 2010 đến 2022 thể hiện sự biến động đáng kể. Lợi nhuận đạt đỉnh vào năm 2010 nhưng sau đó giảm mạnh vào năm 2014. Từ năm 2015 đến 2019, lợi nhuận tăng trở lại và duy trì ở mức ổn định. Tuy nhiên, đến năm 2022, lợi nhuận trước thuế giảm sâu, đạt mức thấp nhất trong toàn bộ giai đoạn theo dõi. Xu hướng này có thể phản ánh những thay đổi trong hoạt động kinh doanh hoặc tác động từ tình hình kinh tế.

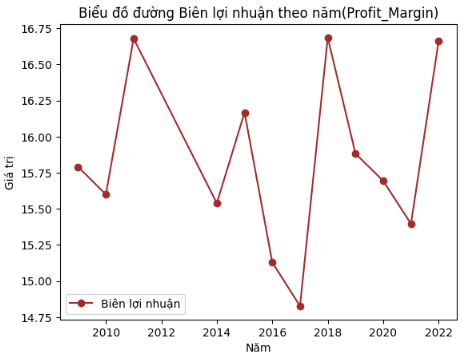
**3.4.3. Trực quan hóa Lợi nhuận ròng theo năm**



Hình 26: Trực quan hóa Lợi nhuận ròng theo năm

Biểu đồ về lợi nhuận ròng từ năm 2009 đến 2022 cho thấy sự biến động đáng kể. Năm 2010, lợi nhuận ròng đạt mức cao nhưng sau đó giảm mạnh vào năm 2014. Từ năm 2015 đến 2019, lợi nhuận phục hồi và duy trì ở mức khá cao, đạt đỉnh vào năm 2019. Tuy nhiên, sau năm 2020, lợi nhuận bắt đầu xu hướng giảm dần, và đến năm 2022 đã giảm mạnh, xuống mức thấp nhất trong toàn bộ giai đoạn. Xu hướng này có thể phản ánh những thay đổi trong chiến lược kinh doanh hoặc tác động của các yếu tố kinh tế.

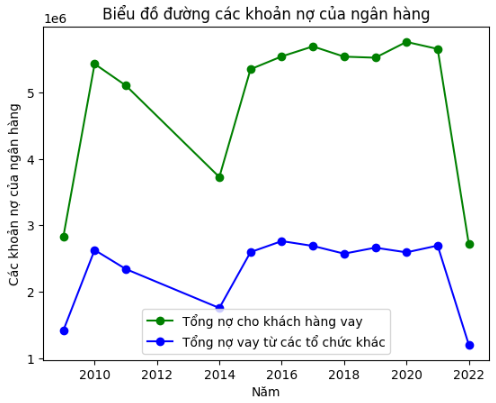
**3.4.4. Trực quan hóa Biên lợi nhuận theo năm**



Hình 26: Trực quan hóa Biên lợi nhuận theo năm

Biểu đồ biên lợi nhuận từ năm 2010 đến 2022 cho thấy sự dao động đáng kể. Các năm 2012 và 2018 ghi nhận mức biên lợi nhuận cao nhất, vượt **16.5%**, trong khi những năm như 2014 và 2016 chứng kiến sự sụt giảm mạnh, đặc biệt năm 2016 xuống dưới **15%**. Đến năm 2022, biên lợi nhuận đã tăng trở lại gần mức đỉnh trước đó. Những thay đổi này có thể do ảnh hưởng của điều kiện kinh tế hoặc điều chỉnh trong chiến lược của ngân hàng.

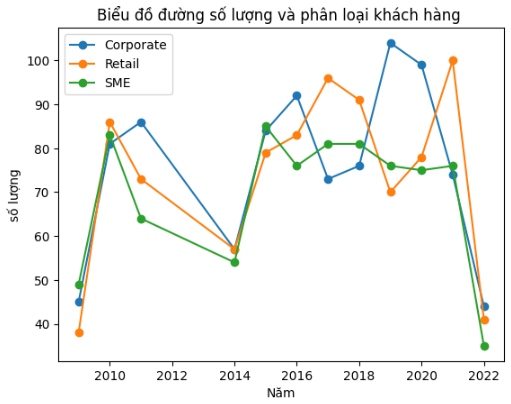
**3.4.5. Trực quan hóa Các khoản vay của ngân hàng**



Hình 28: Trực quan hóa Các khoản vay của ngân hàng

Biểu đồ cho thấy sự biến động của hai loại nợ chính của ngân hàng từ năm 2010 đến 2022: **Tổng nợ cho khách hàng vay** và **Tổng nợ vay từ các tổ chức khác**. Nợ cho khách hàng vay tăng đáng kể từ năm 2010 và đạt mức cao nhất vào năm 2021, sau đó giảm mạnh vào năm 2022. Đối với nợ vay từ các tổ chức khác, sau giai đoạn tăng nhẹ từ 2010 đến 2016, nợ duy trì ổn định cho đến năm 2021 trước khi giảm mạnh vào năm 2022. Xu hướng này có thể phản ánh sự điều chỉnh trong chính sách tài chính của ngân hàng và ảnh hưởng từ môi trường kinh tế.

**3.4.6. Trực quan hóa Số lượng và Phân loại khách hàng**



Hình 29: Trực quan hóa Số lượng và phân loại khách hàng

Biểu đồ thể hiện sự biến động về số lượng khách hàng trong ba nhóm chính: **Corporate** (doanh nghiệp), **Retail** (bán lẻ), và **SME** (doanh nghiệp vừa và nhỏ) từ năm 2010 đến 2022. Số lượng khách hàng ở cả ba nhóm tăng mạnh từ năm 2010 đến 2013, sau đó giảm đột ngột vào năm 2014. Từ năm 2015 đến 2021, số lượng khách hàng trong các nhóm có xu hướng ổn định, với **Retail** và **Corporate** duy trì số lượng cao hơn so với **SME**. Tuy nhiên, năm 2022 ghi nhận sự sụt giảm rõ rệt ở cả ba nhóm, có thể do ảnh hưởng của thay đổi chính sách hoặc điều kiện kinh tế.

# KẾT LUẬN

## Kết quả đạt được.

Dự án của chúng tôi đã đạt được những thành quả quan trọng trong việc phân tích dữ liệu tài chính của ngân hàng VCB, góp phần nâng cao độ chính xác trong dự báo xu hướng tài chính và hỗ trợ các quyết định chiến lược của ngân hàng. Thông qua việc sử dụng các mô hình phân tích hiện đại, ngân hàng có thể nhận diện sớm các biến động về doanh thu, lợi nhuận và nợ vay, từ đó giúp tối ưu hóa hoạt động và giảm thiểu rủi ro tài chính.

Mô hình phân tích không chỉ cung cấp những dữ liệu hữu ích trong việc đưa ra quyết định tài chính, mà còn cho thấy sự ứng dụng hiệu quả của khoa học dữ liệu vào lĩnh vực ngân hàng. Với việc khai thác chi tiết các dữ liệu tài chính, chúng tôi đã phát triển các công cụ phân tích giúp ngân hàng có cái nhìn toàn diện hơn về tình hình kinh doanh. Cách trình bày dữ liệu trực quan hỗ trợ đội ngũ quản lý trong việc nhanh chóng nắm bắt các yếu tố quan trọng và đưa ra hành động kịp thời.

Ngoài ra, mô hình cũng có khả năng dự báo các chỉ số tài chính dựa trên các yếu tố như doanh thu, lợi nhuận và các khoản vay. Điều này giúp ngân hàng đưa ra các biện pháp quản lý tài chính hợp lý hơn và cải thiện khả năng ứng phó với những biến động của thị trường.

Một ưu điểm nổi bật của mô hình là khả năng phân tích và nhận diện các yếu tố quan trọng nhất ảnh hưởng đến hiệu quả tài chính của ngân hàng. Những thông tin này đóng vai trò quan trọng trong việc quản lý rủi ro và tối ưu hóa chiến lược sử dụng nguồn lực. Việc áp dụng mô hình vào thực tế giúp ngân hàng nâng cao khả năng quản lý tài chính, cải thiện hiệu quả trong hoạt động tín dụng và đầu tư, từ đó gia tăng lợi nhuận.

Kết quả của dự án này không chỉ hỗ trợ ngân hàng VCB trong việc phân tích và quản lý tài chính hiệu quả hơn, mà còn góp phần nâng cao khả năng cạnh tranh và hiệu suất hoạt động trên thị trường.

## Hạn chế và hướng phát triển .

**Hạn chế**

Mặc dù dự án đã đạt được một số kết quả tích cực, nhưng cũng tồn tại một số hạn chế đáng kể. Do thời gian thực hiện dự án khá ngắn, chúng em không có đủ điều kiện để tự tổng hợp hoặc thu thập một bộ dữ liệu đủ lớn và chất lượng. Bộ dữ liệu sử dụng trong dự án còn thiếu sự đa dạng và không đủ phong phú, dẫn đến mô hình không thể tổng quát hóa tốt trên các trường hợp mới. Điều này làm giảm độ chính xác của dự đoán và khiến kết quả chưa đạt được sự hoàn thiện mong muốn.

**Hướng phát triển**

Trong tương lai, nếu có thêm thời gian và nguồn lực, chúng em dự định mở rộng và cải thiện chất lượng dữ liệu. Điều này có thể thực hiện bằng cách nghiên cứu và tích hợp thêm dữ liệu từ nhiều nguồn khác nhau, đảm bảo sự phong phú và đa dạng hơn, từ đó cải thiện khả năng dự đoán của mô hình.

Ngoài ra, chúng em cũng hướng đến việc tinh chỉnh mô hình dự đoán. Bằng cách điều chỉnh các tham số và thử nghiệm nhiều cấu hình khác nhau, chúng em sẽ tìm ra phiên bản mô hình tối ưu nhất, đồng thời tránh tình trạng overfitting hoặc underfitting. Những cải tiến này sẽ giúp nâng cao hiệu quả và tính chính xác của mô hình trong việc dự đoán bệnh tiểu đường.

Đó là những hạn chế và hướng phát triển mà nhóm đã suy nghĩ. Rất mong nhận được ý kiến đóng góp từ thầy cô để hoàn thiện dự án hơn nữa.

Chúng em xin chân thành cảm ơn!

# TÀI LIỆU THAM KHẢO

[1] DATA MINING: CONCEPTS AND TECHNIQUES 3RD EDITION

[2] Nguyễn Thanh Tuấn(2009), Deeplearning cơ bản The Legrand Orange Book Template by Mathias Legrand is used : <https://nttuan8.com/sach-deep-learning-co-ban/>

[3] Lecturer in deep learning at the Faculty of Information Technology