BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐẠI HỌC ĐẠI NAM**

Logo, company name

Description automatically generated

**BÀI TẬP LỚN**

**TÊN HỌC PHẦN: HỌC MÁY**

**ĐỀ TÀI: DỰ BÁO NHU CẦU HÀNG TỒN KHO BẰNG CÁCH SỬ DỤNG MÁY HỌC**

**Giáo viên hướng dẫn: Giảng viên Trần Hồng Việt**

**Sinh viên thực hiện:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Stt** | **Họ và tên** | **Mã sv** | **Lớp** |
| 1 | Phạm Đình Nghĩa | 1671020222 | CNTT 16-01 |
| 2 | Nguyễn Đào Phúc Nguyên | 1671020227 | CNTT 16-01 |
| 3 | Nguyễn Tiến Thái | 1671020228 | CNTT 16-01 |
| 4 | Hồ Quang Huy | 1671020137 | CNTT 16-01 |

**Hà Nội, năm 2024**

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC ĐẠI NAM**

Logo, company name

Description automatically generated

**BÀI TẬP LỚN**

**TÊN HỌC PHẦN: HỌC MÁY**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| STT | Họ và Tên | Mã Sinh Viên | Ngày Sinh | Điểm | |
| Bằng Số | Bằng Chữ |
| 1 | Phạm Đình Nghĩa | 1671020222 |  |  |  |
| 2 | Nguyễn Đào Phúc Nguyên | 1671020227 | 13/10/04 |  |  |
| 3 | Nguyễn Tiến Thái | 1671020228 | 15/09/04 |  |  |
| 4 | Hồ Quang Huy | 161020137 |  |  |  |

**ĐỀ TÀI: DỰ BÁO NHU CẦU HÀNG TỒN KHO BẰNG CÁCH SỬ DỤNG MÁY HỌC**

**CÁN BỘ CHẤM THI**

**Hà Nội, năm 2024**

**LỜI MỞ ĐẦU**

Lời nói đầu

Trong thời đại công nghệ 4.0, việc tối ưu hóa quy trình quản lý và vận hành chuỗi cung ứng đã trở thành một yếu tố then chốt quyết định sự thành công của các doanh nghiệp. Một trong những thách thức lớn nhất mà các doanh nghiệp phải đối mặt là việc dự báo chính xác nhu cầu hàng tồn kho, từ đó đảm bảo cân đối giữa nguồn cung và cầu, tránh tình trạng thiếu hụt hoặc tồn đọng hàng hóa. Để giải quyết vấn đề này, việc áp dụng các phương pháp hiện đại như Máy học (Machine Learning) đang dần trở thành xu hướng.

Bài tập lớn này với đề tài "Dự báo nhu cầu hàng tồn kho bằng cách sử dụng máy học" tập trung vào việc nghiên cứu và ứng dụng các kỹ thuật học máy nhằm nâng cao hiệu quả dự báo nhu cầu hàng tồn kho. Qua đó, chúng tôi sẽ sử dụng các mô hình học máy để phân tích dữ liệu lịch sử bán hàng, từ đó dự báo nhu cầu tương lai dựa trên những xu hướng và mẫu hình đã được nhận diện.

Mục tiêu của bài tập lớn không chỉ dừng lại ở việc phát triển một hệ thống dự báo có độ chính xác cao mà còn hướng đến việc đề xuất một quy trình tự động hóa trong việc quản lý tồn kho. Điều này sẽ giúp doanh nghiệp có thể ra quyết định nhanh chóng và chính xác, tối ưu hóa chuỗi cung ứng và giảm thiểu rủi ro phát sinh từ việc dư thừa hoặc thiếu hụt hàng hóa.

**MỤC LỤC**

[**CHƯƠNG 1. TỔNG QUAN VỀ DỰ BÁO HÀNG TỒN KHO VÀ VAI TRÒ CỦA MÁY HỌC** 1](#_Toc179490712)

[1.1 Tổng quan về quản lý hàng tồn kho 1](#_Toc179490713)

[1.2 Các phương pháp truyền thống trong dự báo hàng tồn kho 2](#_Toc179490714)

[1.3 Vai trò của máy học trong dự báo hàng tồn kho 3](#_Toc179490715)

[**CHƯƠNG 2: CÁC MÔ HÌNH MÁY HỌC ỨNG DỤNG TRONG DỰ BÁO HÀNG TỒN KHO** 10](#_Toc179490716)

[2.1 Mô hình hồi quy (Regression Models) 10](#_Toc179490717)

[2.2 Cây quyết định (Decision Trees) 14](#_Toc179490718)

[2.3 Rừng ngẫu nhiên (Random Forest) 18](#_Toc179490719)

[2.4 Học sâu (Deep Learning) 22](#_Toc179490720)

[2.5 Mô hình hồi quy logistic (Logistic Regression) 25](#_Toc179490721)

[2.6 Mô hình ARIMA (AutoRegressive Integrated Moving Average) 27](#_Toc179490722)

[2.7 Học máy nâng cao (Ensemble Learning) 29](#_Toc179490723)

[2.8 Mô hình phân tích chuỗi thời gian (Time Series Analysis) 32](#_Toc179490724)

[**CHƯƠNG 3: PHÂN TÍCH CHƯƠNG TRÌNH DỰ BÁO NHU CẦU HÀNG TỒN KHO BẰNG CÁCH SỬ DỤNG MÁY HỌC** 36](#_Toc179490725)

[3.1. Mô tả chi tiết chương trình 36](#_Toc179490726)

[**TỔNG KẾT CHƯƠNG TRÌNH DỰ BÁO NHU CẦU HÀNG TỒN KHO** 49](#_Toc179490727)

# **TỔNG QUAN VỀ DỰ BÁO HÀNG TỒN KHO VÀ VAI TRÒ CỦA MÁY HỌC**

## 1.1 Tổng quan về quản lý hàng tồn kho

**Quản lý hàng tồn kho** là một phần không thể thiếu trong hoạt động của doanh nghiệp, đặc biệt đối với những doanh nghiệp có hệ thống chuỗi cung ứng phức tạp. Hàng tồn kho bao gồm không chỉ các sản phẩm đã hoàn thiện mà còn các nguyên vật liệu, sản phẩm dở dang hoặc phụ tùng cần thiết cho quá trình sản xuất. Sự cân bằng giữa **cung ứng và nhu cầu** là điều mà doanh nghiệp luôn tìm cách tối ưu, đảm bảo rằng họ có đủ sản phẩm để đáp ứng nhu cầu khách hàng mà không gây dư thừa, từ đó giảm thiểu chi phí lưu kho.

Mục tiêu chính của quản lý hàng tồn kho là **giảm thiểu chi phí tổng thể**, bao gồm:

* **Chi phí lưu trữ**: Bao gồm chi phí thuê kho bãi, bảo dưỡng hàng hóa, và rủi ro về hư hỏng.
* **Chi phí đặt hàng**: Bao gồm chi phí giao dịch, vận chuyển và nhận hàng.
* **Chi phí thiếu hụt**: Khi nhu cầu không được đáp ứng do thiếu hàng, doanh nghiệp sẽ chịu tổn thất từ việc mất doanh thu, giảm sự hài lòng của khách hàng, và ảnh hưởng đến hình ảnh thương hiệu.

Dự báo nhu cầu hàng tồn kho là một khía cạnh then chốt trong quản lý tồn kho. Nó không chỉ giúp doanh nghiệp dự báo được lượng hàng cần thiết cho tương lai mà còn hỗ trợ việc lập kế hoạch sản xuất, lập kế hoạch mua hàng và tối ưu hóa chuỗi cung ứng. Tuy nhiên, dự báo nhu cầu không phải là điều đơn giản do ảnh hưởng của nhiều yếu tố không chắc chắn như:

* **Xu hướng tiêu dùng thay đổi liên tục**.
* **Yếu tố mùa vụ**: Sự biến động theo mùa, chẳng hạn như các mặt hàng quần áo thời trang, thực phẩm, thường có nhu cầu cao trong một khoảng thời gian ngắn.
* **Chương trình khuyến mãi hoặc giảm giá**: Các chiến dịch marketing có thể thúc đẩy nhu cầu tăng đột biến trong một khoảng thời gian ngắn.
* **Sự bất ổn kinh tế**: Các yếu tố như khủng hoảng kinh tế, lạm phát, hoặc thay đổi trong chính sách có thể ảnh hưởng đến hành vi tiêu dùng.

Với các yếu tố phức tạp này, việc **dự báo chính xác** là một nhiệm vụ khó khăn nhưng rất quan trọng, từ đó xuất hiện nhu cầu về các công cụ tiên tiến để giúp doanh nghiệp quản lý tốt hơn nhu cầu hàng tồn kho của mình.

## 1.2 Các phương pháp truyền thống trong dự báo hàng tồn kho

Trong nhiều thập kỷ, các doanh nghiệp đã sử dụng các **phương pháp truyền thống** để dự báo nhu cầu hàng tồn kho. Những phương pháp này thường dựa trên dữ liệu lịch sử và các công cụ thống kê cơ bản để đưa ra dự báo.

* **Phương pháp chuỗi thời gian (Time Series Analysis)**:  
  Phương pháp này dựa trên các dữ liệu bán hàng trong quá khứ để phân tích các xu hướng, chu kỳ hoặc tính mùa vụ. Những kỹ thuật như **trung bình trượt (Moving Average)** và **phân rã mũ (Exponential Smoothing)** là những ví dụ phổ biến. Phương pháp này có ưu điểm là dễ áp dụng, nhưng hạn chế lớn là dựa hoàn toàn vào các mẫu hình trong quá khứ mà không tính đến các yếu tố ngoại vi hay biến động bất ngờ.
* **Phương pháp hồi quy tuyến tính (Linear Regression)**:  
  Đây là một phương pháp cơ bản trong thống kê, mô hình hóa mối quan hệ giữa nhu cầu (biến phụ thuộc) và các yếu tố liên quan khác (biến độc lập) như giá cả, khuyến mãi, hoặc dữ liệu kinh tế. Mặc dù hồi quy tuyến tính có thể mang lại kết quả nhanh chóng và tương đối dễ hiểu, nhưng nó chỉ thích hợp cho các dữ liệu có quan hệ tuyến tính. Đối với các tình huống có sự phi tuyến hay dữ liệu phức tạp, phương pháp này tỏ ra kém hiệu quả.
* **Phương pháp dự báo bằng kinh nghiệm (Judgmental Forecasting)**:  
  Các nhà quản lý có kinh nghiệm lâu năm thường sử dụng sự am hiểu thị trường để dự báo nhu cầu. Tuy nhiên, phương pháp này phụ thuộc rất nhiều vào kinh nghiệm cá nhân và dễ bị sai sót do thành kiến cá nhân, thiếu dữ liệu hoặc sự không ổn định của môi trường kinh doanh.

Các phương pháp này tuy có giá trị nhất định trong các trường hợp đơn giản, nhưng khi đối mặt với dữ liệu phức tạp, các biến động thị trường không thể dự báo trước, hoặc khi khối lượng dữ liệu trở nên quá lớn, chúng không còn đáp ứng được yêu cầu về tính chính xác và linh hoạt. Đây chính là lý do cần thiết phải tìm kiếm những phương pháp tiên tiến hơn.

## 1.3 Vai trò của máy học trong dự báo hàng tồn kho

Máy học (ML) đã thay đổi cách các doanh nghiệp tiếp cận việc dự báo hàng tồn kho, giúp cải thiện độ chính xác và giảm thiểu rủi ro trong quản lý chuỗi cung ứng. Dưới đây là các đặc điểm nổi bật và cách mà máy học vượt trội trong việc dự báo nhu cầu hàng tồn kho, đi sâu hơn vào từng khía cạnh giúp các doanh nghiệp tối ưu hóa quản lý hàng tồn kho.

***1.3.1 Xử lý dữ liệu lớn và đa chiều***

Trong các doanh nghiệp lớn và các tổ chức có chuỗi cung ứng phức tạp, lượng dữ liệu liên quan đến hàng tồn kho là khổng lồ. Các phương pháp dự báo truyền thống thường gặp khó khăn khi phải xử lý lượng dữ liệu lớn và đa chiều này. Máy học cung cấp các giải pháp mạnh mẽ trong việc quản lý và phân tích các khối dữ liệu lớn, từ nhiều nguồn khác nhau.

- *Dữ liệu đa dạng*: Trong quá trình dự báo nhu cầu hàng tồn kho, dữ liệu không chỉ đến từ lịch sử bán hàng mà còn bao gồm dữ liệu về khách hàng, dữ liệu kinh tế, thời tiết, hành vi mua sắm của khách hàng, xu hướng trên mạng xã hội, và các yếu tố chính trị. Máy học có khả năng tổng hợp và xử lý tất cả các dữ liệu này cùng lúc, tạo ra các dự báo có tính chính xác cao hơn nhờ tận dụng được đa dạng các nguồn dữ liệu.

*- Phân tích dữ liệu phi cấu trúc:* Dữ liệu phi cấu trúc, chẳng hạn như nhận xét khách hàng, phản hồi từ mạng xã hội, hoặc các đánh giá từ người dùng, là các thông tin khó phân tích bằng các phương pháp truyền thống. Máy học sử dụng các mô hình xử lý ngôn ngữ tự nhiên (NLP) để phân tích các thông tin này, giúp phát hiện các xu hướng hoặc yếu tố tiềm ẩn có thể ảnh hưởng đến nhu cầu hàng tồn kho.

- *Xử lý dữ liệu thời gian thực:* Trong bối cảnh thị trường luôn biến động, khả năng cập nhật dữ liệu thời gian thực là một yếu tố quan trọng giúp nâng cao tính chính xác của dự báo. Máy học có thể liên tục thu thập, xử lý dữ liệu mới và cập nhật mô hình dự báo theo thời gian thực, giúp doanh nghiệp phản ứng nhanh chóng với các thay đổi bất ngờ trong thị trường.

***1.3.2 Học và thích ứng từ dữ liệu***

Khả năng học hỏi và thích ứng là một trong những ưu điểm mạnh mẽ nhất của máy học trong dự báo nhu cầu hàng tồn kho. Không giống như các mô hình dựa trên quy tắc cố định của phương pháp truyền thống, các thuật toán máy học có khả năng học từ dữ liệu quá khứ và điều chỉnh mô hình để cải thiện dự báo khi có thêm dữ liệu mới.

*- Học từ dữ liệu lịch sử:* Một trong những lợi thế của máy học là khả năng sử dụng dữ liệu lịch sử để học và xây dựng các mô hình dự báo. Các mô hình học máy như Random Forest, Neural Networks, và Gradient Boosting có thể "nhớ" các mối quan hệ phức tạp giữa dữ liệu và phát hiện các mẫu hình mà con người hoặc các mô hình đơn giản có thể bỏ sót.

*- Thích ứng nhanh với sự thay đổi*: Các mô hình máy học có khả năng liên tục học từ dữ liệu mới và cập nhật để phản ánh những thay đổi trong thị trường. Ví dụ, nếu có sự thay đổi đột ngột trong hành vi tiêu dùng do khuyến mãi lớn hoặc một biến động kinh tế, mô hình học máy có thể nhanh chóng điều chỉnh dự báo để đáp ứng với các thay đổi đó. Điều này giúp doanh nghiệp giảm rủi ro về việc thiếu hàng hoặc thừa hàng.

*- Tự điều chỉnh và cải thiện liên tục:* Một trong những đặc điểm nổi bật của máy học là khả năng cải thiện liên tục thông qua các thuật toán học có giám sát hoặc học không giám sát. Các mô hình dự báo được "huấn luyện" bằng cách phân tích dữ liệu cũ, và chúng có thể điều chỉnh các tham số để cải thiện độ chính xác của dự báo. Nhờ vào khả năng này, doanh nghiệp có thể liên tục nâng cao chất lượng quản lý hàng tồn kho mà không cần can thiệp thủ công quá nhiều.

***1.3.3 Nhận diện mẫu hình phi tuyến***

Dự báo hàng tồn kho không phải lúc nào cũng tuân theo các mối quan hệ tuyến tính (linear relationships). Trong nhiều trường hợp, mối quan hệ giữa các yếu tố ảnh hưởng đến nhu cầu là phi tuyến (non-linear), ví dụ như tương quan giữa nhu cầu sản phẩm và các yếu tố ngoại vi như giá cả, khuyến mãi, hoặc xu hướng thị trường. Các phương pháp truyền thống thường gặp khó khăn khi đối mặt với các mối quan hệ phi tuyến này.

*- Mô hình hóa các quan hệ phức tạp:* Các thuật toán máy học như Random Forest, Gradient Boosting, và Neural Networks có khả năng xử lý các mối quan hệ phi tuyến phức tạp. Ví dụ, mô hình học máy có thể phát hiện rằng nhu cầu sản phẩm không tăng đều theo giá cả, mà có thể đạt đỉnh ở một mức giá nhất định rồi giảm khi giá tăng cao hơn nữa. Việc nhận diện các mẫu hình phức tạp này giúp các mô hình máy học đưa ra các dự báo chính xác hơn.

*- Phát hiện các tương tác đa chiều:* Ngoài việc xử lý phi tuyến, máy học có thể phát hiện các mối quan hệ tương tác giữa nhiều yếu tố khác nhau. Chẳng hạn, một yếu tố như thời tiết có thể ảnh hưởng đến nhu cầu sản phẩm theo nhiều cách khác nhau tùy thuộc vào khu vực địa lý và thời điểm trong năm. Máy học giúp doanh nghiệp phát hiện và mô hình hóa các mối quan hệ đa chiều này để tạo ra các dự báo chi tiết và phù hợp hơn với thực tế.

***1.3.4 Dự đoán với các yếu tố ngoại vi và sự biến động***

Thị trường và nhu cầu của khách hàng có thể chịu ảnh hưởng từ rất nhiều yếu tố ngoài tầm kiểm soát của doanh nghiệp, như biến động kinh tế, chính sách thương mại, thời tiết, hay sự kiện xã hội lớn. Các yếu tố này thường không thể dự đoán trước bằng các phương pháp truyền thống. Tuy nhiên, với sự hỗ trợ của máy học, doanh nghiệp có thể đưa vào mô hình các yếu tố ngoại vi này để tăng cường độ chính xác của dự báo.

*- Tích hợp yếu tố kinh tế và xã hội:* Máy học có thể thu thập và tích hợp các dữ liệu kinh tế, xã hội, chính trị vào trong mô hình dự báo. Ví dụ, trong các ngành như thời trang hoặc thực phẩm, sự thay đổi về văn hóa, hoặc xu hướng xã hội có thể tạo ra sự gia tăng đột ngột về nhu cầu cho một số sản phẩm nhất định. Sự biến động của tỷ giá, chính sách thuế, và tình hình kinh tế chung cũng có thể ảnh hưởng đến quyết định mua hàng của người tiêu dùng.

*- Phân tích và dự đoán từ yếu tố thời tiết*: Thời tiết là một trong những yếu tố ngoại vi có thể ảnh hưởng lớn đến nhu cầu hàng tồn kho, đặc biệt trong các ngành công nghiệp như bán lẻ, thực phẩm, và du lịch. Các mô hình máy học có thể tích hợp dữ liệu thời tiết trong thời gian thực và dự đoán ảnh hưởng của nó đến nhu cầu sản phẩm. Chẳng hạn, nhu cầu cho quần áo mùa đông, máy sưởi hoặc đồ uống nóng thường tăng mạnh khi nhiệt độ giảm.

*- Phản ứng với các biến động bất ngờ:* Máy học có thể phân tích nhanh chóng các tình huống bất ngờ như đại dịch, thiên tai, hoặc biến động thị trường. Các thuật toán máy học không chỉ có khả năng điều chỉnh mô hình dự báo mà còn có thể giúp doanh nghiệp tìm ra các phương án thay thế để tối ưu hóa hoạt động kinh doanh trong các tình huống khẩn cấp.

***1.3.5 Tự động hóa quá trình dự báo và giảm phụ thuộc vào chuyên gia***

Một trong những lợi ích lớn của việc áp dụng máy học trong dự báo nhu cầu hàng tồn kho là khả năng tự động hóa quá trình dự báo mà không cần phụ thuộc quá nhiều vào chuyên gia. Điều này đặc biệt hữu ích khi doanh nghiệp mở rộng quy mô, và việc dự báo thủ công trở nên khó khăn.

*- Tự động hóa hoàn toàn*: Với các phương pháp máy học, doanh nghiệp có thể thiết lập các hệ thống tự động thu thập, phân tích dữ liệu và cập nhật dự báo theo thời gian thực. Các mô hình học máy có thể tự động tối ưu hóa các tham số, giảm thiểu nhu cầu can thiệp từ con người và cho phép các nhóm quản lý hàng tồn kho tập trung vào việc ra quyết định chiến lược thay vì thực hiện các công việc dự báo thủ công.

*- Giảm sai lệch do yếu tố con người:* Khi dự báo dựa trên sự đánh giá của các chuyên gia, yếu tố cảm tính hoặc kinh nghiệm cá nhân có thể ảnh hưởng đến kết quả dự báo. Điều này có thể gây ra các lỗi dự báo không mong muốn. Máy học loại bỏ sự ảnh hưởng của yếu tố cảm tính và đảm bảo các quyết định dựa trên dữ liệu khách quan.

***1.3.6 Khả năng dự báo ngắn hạn và dài hạn***

Máy học không chỉ có khả năng dự báo ngắn hạn mà còn có khả năng xử lý các dự báo dài hạn một cách chính xác. Điều này rất quan trọng trong quản lý hàng tồn kho, nơi mà doanh nghiệp cần phải có cái nhìn tổng quan về nhu cầu trong nhiều khoảng thời gian khác nhau để tối ưu hóa quy trình cung ứng.

*- Dự báo ngắn hạn*: Các mô hình máy học như Hồi quy tuyến tính hoặc Mạng Nơ-ron nhân tạo (Artificial Neural Networks) có thể dễ dàng phát hiện các xu hướng ngắn hạn trong dữ liệu. Điều này giúp dự đoán nhu cầu trong các giai đoạn ngắn hạn như tuần, tháng hoặc quý, giúp doanh nghiệp tối ưu hóa lượng hàng tồn kho trong thời gian gần.

*- Dự báo dài hạn*: Đối với các mô hình chuỗi thời gian dài hơn, các thuật toán như LSTM (Long Short-Term Memory) là rất hiệu quả. LSTM giúp xử lý các dự báo liên quan đến xu hướng dài hạn, chẳng hạn như kế hoạch sản xuất hàng năm hoặc chiến lược mở rộng thị trường.

***1.3.7 Phát hiện xu hướng ẩn và khả năng dự đoán sự bất thường***

Một ưu điểm nổi bật khác của các mô hình máy học là khả năng phát hiện xu hướng ẩn trong dữ liệu và dự đoán các sự kiện bất thường. Đây là điều mà các phương pháp truyền thống thường gặp khó khăn vì chúng thường chỉ dựa vào các giả định hoặc mô hình tuyến tính đơn giản.

*- Phát hiện xu hướng ẩn:* Máy học không chỉ dựa vào các yếu tố rõ ràng, mà còn có khả năng khai phá và phát hiện ra các xu hướng ẩn mà con người hoặc các mô hình dự đoán tuyến tính có thể bỏ sót. Ví dụ, một sản phẩm có thể đột ngột trở nên phổ biến do sự xuất hiện của một xu hướng xã hội hoặc văn hóa mà trước đó không thể lường trước. Các mô hình như Gradient Boosting hay Random Forest có thể phân tích và nhận ra những xu hướng tiềm ẩn này, giúp doanh nghiệp có sự chuẩn bị trước.

*- Dự đoán sự bất thường:* Những sự kiện bất thường trong dữ liệu, như tăng vọt bất ngờ về nhu cầu hay các tình trạng khan hàng ngắn hạn, có thể làm rối loạn kế hoạch hàng tồn kho. Các thuật toán như Isolation Forest hay k-Nearest Neighbors (k-NN) có khả năng nhận diện các mẫu hình bất thường trong dữ liệu và dự báo trước khi chúng xảy ra. Điều này giúp doanh nghiệp phản ứng nhanh chóng, giảm thiểu tác động tiêu cực và duy trì tính ổn định trong quản lý hàng tồn kho.

*- Giảm thiểu rủi ro:* Máy học giúp phát hiện sớm các yếu tố có thể dẫn đến rủi ro, chẳng hạn như việc dự báo sai lầm về nhu cầu hoặc sự thay đổi đột ngột trong hành vi tiêu dùng. Nhờ vào khả năng phân tích dữ liệu liên tục, máy học có thể phát hiện các dấu hiệu cảnh báo sớm và đưa ra khuyến nghị điều chỉnh trước khi vấn đề trở nên nghiêm trọng.

***1.3.8 Tối ưu hóa quy trình chuỗi cung ứng***

Dự báo chính xác không chỉ giúp quản lý tốt hơn hàng tồn kho mà còn giúp tối ưu hóa toàn bộ chuỗi cung ứng. Điều này bao gồm việc lập kế hoạch sản xuất, vận chuyển, và phân phối hàng hóa một cách hiệu quả. Máy học đóng vai trò quan trọng trong việc này bằng cách cung cấp các dự báo chi tiết và đáng tin cậy.

*- Tối ưu hóa vận chuyển và lưu trữ*: Dự báo chính xác giúp doanh nghiệp tối ưu hóa quy trình vận chuyển và lưu trữ hàng hóa. Khi nhu cầu được dự báo chính xác, doanh nghiệp có thể lên kế hoạch vận chuyển hàng hóa từ nhà máy hoặc nhà cung cấp đến các điểm phân phối một cách hiệu quả hơn. Điều này giúp tiết kiệm chi phí vận chuyển và giảm thiểu tình trạng hàng hóa bị tồn đọng quá lâu tại kho.

*- Quản lý sản xuất hiệu quả:* Với các dự báo chính xác về nhu cầu hàng tồn kho, doanh nghiệp có thể lên kế hoạch sản xuất một cách chính xác, đảm bảo rằng nguồn lực được sử dụng một cách hiệu quả mà không bị lãng phí. Điều này giúp giảm chi phí sản xuất, tránh tình trạng dư thừa nguyên vật liệu hoặc thiếu hàng thành phẩm để cung cấp cho thị trường.

*- Phân bổ hàng hóa thông minh:* Máy học không chỉ giúp dự báo nhu cầu tổng thể mà còn hỗ trợ phân bổ hàng hóa một cách thông minh đến các khu vực hoặc cửa hàng khác nhau dựa trên nhu cầu của từng khu vực. Chẳng hạn, các mô hình dự báo có thể phân tích dữ liệu tiêu dùng từ các cửa hàng trong khu vực thành phố và vùng nông thôn, từ đó tối ưu hóa lượng hàng phân phối để tránh thiếu hụt hoặc dư thừa không cần thiết.

***1.3.9 Tăng cường khả năng cạnh tranh và đưa ra quyết định chiến lược***

Khả năng dự báo chính xác nhu cầu hàng tồn kho không chỉ giúp doanh nghiệp tối ưu hóa hoạt động hàng ngày mà còn mang lại lợi thế cạnh tranh quan trọng trên thị trường. Máy học giúp doanh nghiệp ra quyết định chiến lược, giúp họ chuẩn bị tốt hơn cho những thay đổi trong thị trường và phản ứng nhanh với các cơ hội mới.

*- Nâng cao năng lực cạnh tranh:* Trong thị trường cạnh tranh gay gắt, việc dự báo nhu cầu chính xác giúp doanh nghiệp tránh được tình trạng thiếu hụt hoặc dư thừa hàng hóa, từ đó giảm chi phí và tối đa hóa lợi nhuận. Đồng thời, với khả năng dự đoán xu hướng tiêu dùng, doanh nghiệp có thể đi trước đối thủ bằng cách đưa ra các sản phẩm hoặc chiến lược tiếp cận thị trường phù hợp hơn, từ đó giành được thị phần lớn hơn.

*- Ra quyết định chiến lược thông minh hơn*: Các doanh nghiệp không chỉ cần dự báo để quản lý hàng tồn kho mà còn cần các dự báo chính xác để định hình các quyết định chiến lược dài hạn. Máy học cung cấp thông tin dự báo không chỉ cho các hoạt động hàng ngày mà còn giúp doanh nghiệp lập kế hoạch chiến lược trong dài hạn, chẳng hạn như mở rộng thị trường, phát triển sản phẩm mới, hoặc thay đổi chiến lược tiếp thị.

*- Phát hiện cơ hội mới:* Nhờ vào khả năng phân tích dữ liệu lớn, máy học giúp doanh nghiệp nhận diện được các cơ hội mới trên thị trường, chẳng hạn như sự gia tăng nhu cầu cho một dòng sản phẩm cụ thể hoặc xu hướng tiêu dùng mới xuất hiện. Việc nhận biết các cơ hội này sớm cho phép doanh nghiệp điều chỉnh chiến lược nhanh chóng, từ đó nắm bắt được lợi thế thị trường trước đối thủ cạnh tranh.

# **CHƯƠNG 2: CÁC MÔ HÌNH MÁY HỌC ỨNG DỤNG TRONG DỰ BÁO HÀNG TỒN KHO**

Dự báo nhu cầu hàng tồn kho đóng vai trò quan trọng trong quản lý chuỗi cung ứng và quyết định về sản xuất, lưu kho, và phân phối. Việc áp dụng các mô hình máy học không chỉ giúp nâng cao độ chính xác của dự báo mà còn giúp doanh nghiệp tối ưu hóa quy trình và giảm thiểu chi phí. Chương này sẽ trình bày các mô hình máy học phổ biến cùng với ưu điểm và ứng dụng của chúng trong dự báo hàng tồn kho.

## 2.1 Mô hình hồi quy (Regression Models)

Mô hình hồi quy là một trong những phương pháp thống kê và máy học quan trọng nhất, thường được sử dụng để phân tích và dự đoán các biến. Mục tiêu chính của hồi quy là tìm ra mối quan hệ giữa một biến phụ thuộc (biến mà chúng ta muốn dự đoán, như nhu cầu hàng tồn kho) và một hoặc nhiều biến độc lập (các yếu tố có thể ảnh hưởng đến biến phụ thuộc). Điều này có ý nghĩa sâu sắc trong việc tối ưu hóa quản lý hàng tồn kho, giúp các doanh nghiệp đưa ra quyết định dựa trên dữ liệu và thông tin chính xác.

***2.1.1 Khái niệm và nguyên lý hoạt động của hồi quy***

Mô hình hồi quy nhằm mục đích mô hình hóa mối quan hệ giữa biến phụ thuộc và các biến độc lập thông qua các tham số. Khi một mô hình hồi quy được xây dựng, nó sẽ cho phép chúng ta ước lượng nhu cầu hàng tồn kho dựa trên một số yếu tố đầu vào. Để đạt được điều này, mô hình cần có dữ liệu lịch sử về nhu cầu hàng hóa và các yếu tố ảnh hưởng liên quan.

**Nguyên lý hoạt động của hồi quy** dựa trên việc tối ưu hóa một hàm mục tiêu, thường là tổng bình phương sai số giữa giá trị dự đoán và giá trị thực tế. Bằng cách điều chỉnh các tham số của mô hình, chúng ta cố gắng tìm ra mô hình tốt nhất có thể mô tả dữ liệu. Sau đó, mô hình này có thể được áp dụng để dự đoán giá trị trong tương lai.

***2.1.2 Phân loại mô hình hồi quy***

Mô hình hồi quy có thể được phân loại thành nhiều loại khác nhau dựa trên số lượng biến độc lập và hình thức của mối quan hệ giữa chúng:

* **Hồi quy tuyến tính đơn giản**: Là mô hình đơn giản nhất, trong đó chỉ có một biến độc lập. Mô hình này giả định rằng mối quan hệ giữa biến phụ thuộc và biến độc lập là tuyến tính, tức là biến phụ thuộc sẽ thay đổi một cách đồng nhất khi biến độc lập thay đổi. Ví dụ, khi phân tích ảnh hưởng của giá cả đến nhu cầu hàng hóa, chúng ta có thể sử dụng hồi quy tuyến tính đơn giản để dự đoán nhu cầu dựa trên giá cả.
* **Hồi quy tuyến tính bội (đa biến)**: Là mô hình có nhiều biến độc lập. Trong thực tế, nhu cầu hàng hóa thường bị ảnh hưởng bởi nhiều yếu tố khác nhau. Do đó, hồi quy tuyến tính bội thường được sử dụng để dự đoán chính xác hơn. Ví dụ, khi dự đoán nhu cầu hàng tồn kho, mô hình có thể xem xét nhiều yếu tố như giá cả, khuyến mãi, mùa vụ, và xu hướng tiêu dùng. Hồi quy tuyến tính bội cho phép người phân tích nắm bắt được ảnh hưởng tổng hợp của nhiều yếu tố đến nhu cầu hàng hóa.
* **Hồi quy phi tuyến**: Trong trường hợp mối quan hệ giữa biến phụ thuộc và biến độc lập không phải là tuyến tính, chúng ta có thể sử dụng các mô hình hồi quy phi tuyến. Những mô hình này có thể bao gồm hồi quy bậc hai hoặc hồi quy với các hàm phi tuyến khác. Ví dụ, nếu sự thay đổi trong giá cả có ảnh hưởng không đồng nhất đến nhu cầu hàng hóa ở các mức giá khác nhau, mô hình hồi quy phi tuyến có thể cung cấp dự đoán chính xác hơn.

***2.1.3 Quy trình xây dựng mô hình hồi quy***

Quy trình xây dựng mô hình hồi quy bao gồm một số bước chính:

1. **Thu thập dữ liệu**: Bước đầu tiên là thu thập dữ liệu lịch sử về biến phụ thuộc và các biến độc lập. Dữ liệu này có thể được lấy từ các hệ thống quản lý hàng tồn kho, báo cáo doanh thu, và các nguồn dữ liệu khác. Độ lớn và chất lượng của tập dữ liệu có thể ảnh hưởng lớn đến kết quả cuối cùng của mô hình.
2. **Tiền xử lý dữ liệu**: Trước khi xây dựng mô hình, cần phải xử lý dữ liệu để đảm bảo rằng nó đầy đủ và chính xác. Bước này bao gồm việc loại bỏ các giá trị thiếu, xử lý các giá trị ngoại lai (outliers) và chuẩn hóa dữ liệu nếu cần thiết. Việc tiền xử lý dữ liệu là rất quan trọng, vì dữ liệu không sạch có thể dẫn đến các kết quả không chính xác và sai lệch.
3. **Phân tích và lựa chọn biến**: Sau khi xử lý dữ liệu, bước tiếp theo là phân tích các biến độc lập để xác định những biến nào có ảnh hưởng lớn nhất đến biến phụ thuộc. Việc lựa chọn đúng các biến sẽ giúp cải thiện độ chính xác của mô hình. Các kỹ thuật như phân tích tương quan, hồi quy từng bước, và kiểm định giả thuyết có thể được sử dụng để lựa chọn biến.
4. **Xây dựng mô hình**: Sử dụng dữ liệu đã chuẩn bị, mô hình hồi quy sẽ được xây dựng bằng cách tối ưu hóa các tham số để giảm thiểu sai số giữa giá trị dự đoán và giá trị thực tế. Trong bước này, các kỹ thuật thống kê và máy học sẽ được áp dụng để tìm ra mô hình tốt nhất.
5. **Đánh giá mô hình**: Sau khi mô hình được xây dựng, cần đánh giá độ chính xác của nó bằng cách sử dụng các chỉ số thống kê như R-squared (hệ số xác định), Root Mean Square Error (RMSE), và Mean Absolute Error (MAE). Bước này sẽ giúp xác định xem mô hình có hoạt động hiệu quả hay không và cần điều chỉnh gì không. Đánh giá mô hình cũng có thể bao gồm việc phân tích dư lượng để kiểm tra tính hợp lý của giả định hồi quy.
6. **Triển khai mô hình**: Nếu mô hình đạt yêu cầu về độ chính xác, nó sẽ được triển khai vào thực tế để phục vụ cho việc dự đoán nhu cầu hàng tồn kho trong tương lai. Mô hình có thể được tích hợp vào các hệ thống quản lý chuỗi cung ứng hoặc phần mềm quản lý hàng tồn kho để tự động hóa quá trình dự đoán.

***2.1.4 Ứng dụng của mô hình hồi quy trong dự báo hàng tồn kho***

Mô hình hồi quy đã được áp dụng rộng rãi trong dự báo nhu cầu hàng tồn kho tại nhiều lĩnh vực khác nhau. Một số ứng dụng chính bao gồm:

* **Dự đoán nhu cầu theo mùa**: Nhu cầu hàng hóa thường có tính chất theo mùa, và mô hình hồi quy có thể giúp dự đoán các đỉnh và đáy trong nhu cầu. Ví dụ, trong ngành bán lẻ, nhu cầu về đồ trang trí trong dịp lễ thường tăng cao, trong khi nhu cầu về hàng hóa khác có thể giảm. Hồi quy cho phép doanh nghiệp lập kế hoạch sản xuất và lưu kho một cách hiệu quả. Việc dự đoán này không chỉ giúp các nhà quản lý tránh tình trạng thiếu hàng mà còn giảm thiểu chi phí lưu kho.
* **Phân tích tác động của các yếu tố bên ngoài**: Hồi quy giúp doanh nghiệp phân tích các yếu tố bên ngoài như biến động giá cả, các chiến dịch quảng cáo và các sự kiện đặc biệt đến nhu cầu hàng hóa. Bằng cách xác định rõ các yếu tố này, doanh nghiệp có thể điều chỉnh chiến lược kinh doanh và tối ưu hóa quản lý hàng tồn kho. Ví dụ, nếu một chiến dịch khuyến mãi thành công, mô hình hồi quy có thể giúp dự đoán nhu cầu tăng cao trong thời gian khuyến mãi.
* **Quản lý hàng tồn kho tối ưu**: Thông qua việc dự đoán chính xác nhu cầu, mô hình hồi quy giúp doanh nghiệp quản lý hàng tồn kho một cách hiệu quả hơn. Doanh nghiệp có thể giảm thiểu tình trạng thừa hàng hóa hoặc thiếu hụt hàng hóa, từ đó tiết kiệm chi phí và nâng cao sự hài lòng của khách hàng. Việc tối ưu hóa hàng tồn kho không chỉ giúp giảm chi phí mà còn tạo ra lợi thế cạnh tranh cho doanh nghiệp.

***2.1.5 Hạn chế của mô hình hồi quy***

Mặc dù mô hình hồi quy là một công cụ mạnh mẽ trong dự báo nhu cầu hàng tồn kho, nó cũng có một số hạn chế nhất định:

* **Giả định về mối quan hệ tuyến tính**: Hồi quy tuyến tính giả định rằng có một mối quan hệ tuyến tính giữa biến phụ thuộc và các biến độc lập. Nếu mối quan hệ thực tế là phi tuyến, mô hình này có thể không cung cấp dự đoán chính xác. Để giải quyết vấn đề này, các nhà phân tích có thể sử dụng hồi quy phi tuyến hoặc các mô hình khác như cây quyết định hoặc mạng nơ-ron.
* **Nhạy cảm với giá trị ngoại lai**: Mô hình hồi quy có thể bị ảnh hưởng đáng kể bởi các giá trị ngoại lai. Các giá trị này có thể làm sai lệch các tham số của mô hình và dẫn đến dự đoán không chính xác. Việc phát hiện và xử lý các giá trị ngoại lai là rất quan trọng để đảm bảo tính chính xác của mô hình.
* **Không thể nắm bắt các yếu tố phức tạp**: Trong thực tế, nhu cầu hàng hóa thường bị ảnh hưởng bởi nhiều yếu tố phức tạp, và mô hình hồi quy có thể không đủ khả năng để nắm bắt hết tất cả các mối quan hệ này. Các yếu tố như tâm lý tiêu dùng, xu hướng thị trường, và các sự kiện không thể đoán trước có thể ảnh hưởng đến nhu cầu mà mô hình hồi quy không thể phản ánh đầy đủ

## 2.2 Cây quyết định (Decision Trees)

Cây quyết định là một trong những phương pháp phân tích dữ liệu và máy học phổ biến, đặc biệt trong lĩnh vực dự đoán và phân loại. Với cấu trúc giống như một cây, mỗi nút trong cây đại diện cho một thuộc tính (hay biến) cụ thể, trong khi các nhánh thể hiện các giá trị của thuộc tính đó, và các lá (nút cuối cùng) đại diện cho quyết định hoặc kết quả cuối cùng. Cây quyết định là một công cụ mạnh mẽ giúp doanh nghiệp trong việc dự đoán nhu cầu hàng tồn kho dựa trên nhiều yếu tố khác nhau.

***2.2.1 Khái niệm và nguyên lý hoạt động của cây quyết định***

Cây quyết định hoạt động dựa trên nguyên lý phân chia dữ liệu thành các nhóm con dựa trên các thuộc tính của dữ liệu. Quy trình xây dựng cây quyết định bao gồm việc chọn thuộc tính tốt nhất để phân chia dữ liệu tại mỗi nút, tiếp tục phân chia cho đến khi đạt được một tiêu chuẩn dừng nhất định. Mục tiêu là giảm thiểu độ hỗn loạn trong từng nhóm con, được đo lường bằng các chỉ số như độ tinh khiết thông tin (Information Gain) hoặc độ Gini (Gini Index).

Mỗi nhánh trong cây quyết định đại diện cho một quyết định dựa trên thuộc tính cụ thể, giúp người dùng dễ dàng theo dõi và hiểu quá trình ra quyết định. Khi sử dụng cây quyết định để dự đoán nhu cầu hàng tồn kho, người dùng có thể bắt đầu từ nút gốc (nút đầu tiên) và lần lượt đi qua các nhánh cho đến khi đến được nút lá, từ đó đưa ra dự đoán về nhu cầu hàng hóa.

***2.2.2 Phân loại cây quyết định***

Cây quyết định có thể được phân loại thành hai loại chính:

* **Cây quyết định phân loại**: Được sử dụng khi biến phụ thuộc là rời rạc (ví dụ: cao, trung bình, thấp). Cây quyết định phân loại thường áp dụng trong các bài toán phân loại, nơi mà mục tiêu là gán nhãn cho một đối tượng dựa trên các đặc tính của nó.
* **Cây quyết định hồi quy**: Được sử dụng khi biến phụ thuộc là liên tục (ví dụ: giá trị doanh thu, nhu cầu hàng tồn kho). Cây quyết định hồi quy giúp ước lượng giá trị của biến phụ thuộc dựa trên các giá trị của biến độc lập.

***2.2.3 Quy trình xây dựng cây quyết định***

Quy trình xây dựng cây quyết định bao gồm các bước sau:

1. **Thu thập dữ liệu**: Bước đầu tiên trong quy trình xây dựng cây quyết định là thu thập dữ liệu lịch sử liên quan đến nhu cầu hàng tồn kho và các yếu tố ảnh hưởng đến nó. Dữ liệu này có thể bao gồm thông tin về doanh thu hàng tháng, giá cả, mùa vụ, các chiến dịch quảng cáo, và các yếu tố bên ngoài khác như kinh tế vĩ mô.
2. **Tiền xử lý dữ liệu**: Dữ liệu cần được làm sạch và chuẩn bị trước khi sử dụng cho việc xây dựng cây quyết định. Việc này bao gồm:
   * **Xử lý giá trị thiếu**: Các giá trị thiếu có thể làm giảm độ chính xác của mô hình. Chúng có thể được xử lý bằng cách loại bỏ các mẫu có giá trị thiếu hoặc điền giá trị trung bình/giá trị phổ biến của biến.
   * **Xử lý các giá trị ngoại lai**: Các giá trị ngoại lai có thể gây ảnh hưởng lớn đến kết quả của cây quyết định. Việc phát hiện và loại bỏ chúng có thể cải thiện tính chính xác của mô hình.
   * **Chuyển đổi các biến phân loại thành dạng số**: Cây quyết định yêu cầu dữ liệu phải ở dạng số. Do đó, các biến phân loại có thể được chuyển đổi thành các biến nhị phân hoặc sử dụng mã hóa một nóng (one-hot encoding).
3. **Chọn thuộc tính tốt nhất**: Tại mỗi nút trong cây quyết định, cần phải chọn thuộc tính tốt nhất để phân chia dữ liệu. Các tiêu chí phổ biến để đánh giá thuộc tính bao gồm:
   * **Độ tinh khiết thông tin (Information Gain)**: Đo lường mức độ giảm thiểu độ không chắc chắn (entropy) trong dữ liệu sau khi phân chia. Thuộc tính có độ tinh khiết thông tin cao sẽ được chọn.
   * **Độ Gini (Gini Index)**: Là một chỉ số khác để đo lường độ phân tán của dữ liệu. Nó cho biết xác suất mà hai mẫu ngẫu nhiên sẽ thuộc về các lớp khác nhau. Một độ Gini thấp cho thấy dữ liệu được phân chia đồng nhất hơn.
4. **Tạo các nhánh**: Sau khi chọn thuộc tính tốt nhất, cây quyết định sẽ được chia thành các nhánh dựa trên các giá trị của thuộc tính đó. Quá trình này tiếp tục lặp lại cho từng nhánh con cho đến khi đạt được tiêu chuẩn dừng, chẳng hạn như số lượng mẫu tối thiểu trong một nhánh hoặc độ sâu tối đa của cây.
5. **Xây dựng và tối ưu hóa cây**: Cây quyết định có thể trở nên rất phức tạp nếu không được tối ưu hóa, dẫn đến hiện tượng quá khớp (overfitting). Để khắc phục điều này, các phương pháp như cắt tỉa (pruning) có thể được áp dụng. Cắt tỉa giúp loại bỏ các nhánh không cần thiết, từ đó làm tăng khả năng tổng quát của mô hình. Các phương pháp cắt tỉa có thể bao gồm:
   * **Cắt tỉa theo độ sâu**: Giới hạn độ sâu tối đa của cây để giảm thiểu độ phức tạp.
   * **Cắt tỉa sau khi xây dựng**: Đánh giá độ chính xác của mô hình trên tập kiểm tra và loại bỏ các nhánh không cải thiện độ chính xác.
6. **Đánh giá mô hình**: Để đánh giá hiệu quả của cây quyết định, các chỉ số như độ chính xác (accuracy), độ nhạy (sensitivity), và độ đặc hiệu (specificity) có thể được sử dụng. Việc kiểm tra độ chính xác của mô hình thường được thực hiện bằng cách sử dụng tập dữ liệu kiểm tra tách biệt với tập dữ liệu huấn luyện. Điều này cho phép đánh giá khả năng tổng quát của mô hình.

***2.2.4 Ưu điểm của cây quyết định***

Cây quyết định mang lại nhiều ưu điểm nổi bật, bao gồm:

* **Dễ hiểu và dễ giải thích**: Cấu trúc của cây quyết định rất dễ hình dung và dễ hiểu, giúp người dùng nắm bắt được cách thức ra quyết định một cách rõ ràng. Điều này rất quan trọng trong việc giải thích các quyết định cho các bên liên quan, đặc biệt trong các tình huống cần báo cáo cho cấp quản lý hoặc các bên liên quan khác.
* **Không yêu cầu giả định về phân phối**: Cây quyết định không yêu cầu dữ liệu phải tuân theo một phân phối cụ thể, điều này làm cho nó linh hoạt hơn so với một số mô hình thống kê khác như hồi quy tuyến tính.
* **Có thể xử lý dữ liệu phi tuyến**: Cây quyết định có khả năng xử lý các mối quan hệ phi tuyến giữa biến phụ thuộc và các biến độc lập mà không cần phải chuyển đổi dữ liệu. Điều này cho phép mô hình phản ánh chính xác hơn các mối quan hệ phức tạp trong dữ liệu.
* **Tự động lựa chọn thuộc tính**: Cây quyết định có khả năng tự động chọn thuộc tính tốt nhất cho từng bước phân chia, giúp tiết kiệm thời gian trong quá trình phân tích dữ liệu. Điều này có thể là một lợi thế lớn trong các bài toán phức tạp, nơi mà số lượng biến có thể rất lớn.
* **Có thể xử lý cả dữ liệu số và phân loại**: Cây quyết định có thể làm việc với cả biến số liên tục và biến phân loại, cho phép người phân tích linh hoạt hơn trong việc xử lý dữ liệu.

***2.2.5 Hạn chế của cây quyết định***

Mặc dù cây quyết định có nhiều ưu điểm, nhưng cũng tồn tại một số hạn chế nhất định:

* **Quá khớp (overfitting)**: Cây quyết định có xu hướng tạo ra các mô hình phức tạp nếu không được tối ưu hóa, dẫn đến việc nó học quá mức từ dữ liệu huấn luyện và không tổng quát tốt cho dữ liệu mới. Quá khớp có thể dẫn đến những quyết định không chính xác trong các tình huống thực tế.
* **Nhạy cảm với dữ liệu không cân bằng**: Khi dữ liệu không cân bằng (ví dụ: số lượng các lớp khác nhau không đồng đều), cây quyết định có thể đưa ra các dự đoán sai lệch. Điều này đặc biệt quan trọng trong các bài toán phân loại, nơi mà một lớp có thể chiếm ưu thế lớn hơn lớp khác.
* **Thiếu tính ổn định**: Một thay đổi nhỏ trong dữ liệu đầu vào có thể dẫn đến sự thay đổi lớn trong cấu trúc của cây quyết định, gây ra sự thiếu ổn định. Do đó, việc đảm bảo tính ổn định của mô hình trong môi trường thực tế là một thách thức.
* **Khó khăn trong việc xử lý các mối quan hệ tương tác phức tạp**: Mặc dù cây quyết định có thể xử lý dữ liệu phi tuyến, nhưng nó có thể gặp khó khăn trong việc nắm bắt các mối quan hệ tương tác phức tạp giữa nhiều thuộc tính. Đôi khi, các mô hình phức tạp hơn như mạng nơ-ron có thể cần thiết để nắm bắt đầy đủ các mối quan hệ này.

***2.2.6 Ứng dụng của cây quyết định trong dự báo hàng tồn kho***

Cây quyết định có thể được áp dụng hiệu quả trong nhiều lĩnh vực khác nhau, đặc biệt là trong việc dự đoán nhu cầu hàng tồn kho:

* **Dự đoán nhu cầu**: Cây quyết định có thể giúp doanh nghiệp phân tích các yếu tố ảnh hưởng đến nhu cầu hàng hóa, từ đó dự đoán nhu cầu trong tương lai. Ví dụ, một mô hình cây quyết định có thể sử dụng các yếu tố như mùa vụ, giá cả, và quảng cáo để dự đoán nhu cầu hàng hóa trong dịp lễ.
* **Phân tích tình huống**: Cây quyết định có thể được sử dụng để phân tích các tình huống khác nhau và đưa ra quyết định phù hợp. Ví dụ, doanh nghiệp có thể sử dụng cây quyết định để xác định xem có nên đặt hàng thêm hay không dựa trên các yếu tố như tồn kho hiện tại và xu hướng nhu cầu.
* **Tối ưu hóa chiến lược tiếp thị**: Doanh nghiệp có thể sử dụng cây quyết định để phân tích tác động của các chiến dịch tiếp thị đến nhu cầu hàng hóa. Từ đó, họ có thể tối ưu hóa các chiến lược tiếp thị và quảng cáo để đạt được hiệu quả tốt nhất. Điều này giúp doanh nghiệp không chỉ tăng doanh thu mà còn tối ưu hóa nguồn lực.
* **Quản lý rủi ro**: Cây quyết định cũng có thể hỗ trợ trong việc quản lý rủi ro liên quan đến nhu cầu hàng tồn kho. Bằng cách phân tích các kịch bản khác nhau và xác định các yếu tố ảnh hưởng đến rủi ro, doanh nghiệp có thể đưa ra các chiến lược ứng phó hiệu quả hơn.
* **Dự đoán xu hướng thị trường**: Cây quyết định có thể được sử dụng để phân tích các xu hướng thị trường và dự đoán nhu cầu trong tương lai dựa trên các yếu tố kinh tế và xã hội. Điều này giúp doanh nghiệp có cái nhìn tổng quan hơn về thị trường và đưa ra các quyết định chiến lược hợp lý.

## 2.3 Rừng ngẫu nhiên (Random Forest)

Rừng ngẫu nhiên (Random Forest) là một trong những phương pháp máy học mạnh mẽ và phổ biến, được phát triển từ mô hình cây quyết định. Rừng ngẫu nhiên là một kỹ thuật học tập tập hợp (ensemble learning), trong đó nhiều cây quyết định được kết hợp với nhau để cải thiện độ chính xác và khả năng tổng quát của mô hình. Phương pháp này đã trở thành một trong những công cụ quan trọng trong việc dự đoán và phân tích dữ liệu, đặc biệt trong các ứng dụng như dự báo nhu cầu hàng tồn kho.

***2.3.1 Khái niệm và nguyên lý hoạt động***

Rừng ngẫu nhiên hoạt động dựa trên nguyên tắc “học từ những lỗi” (learning from mistakes) và việc kết hợp nhiều cây quyết định độc lập để đưa ra một dự đoán cuối cùng. Bằng cách sử dụng phương pháp lấy mẫu ngẫu nhiên (bagging), rừng ngẫu nhiên tạo ra nhiều cây quyết định từ các tập con của dữ liệu huấn luyện, sau đó kết hợp kết quả của các cây này để đưa ra dự đoán cuối cùng.

* **Lấy mẫu ngẫu nhiên (Bootstrap Sampling)**: Rừng ngẫu nhiên sử dụng kỹ thuật lấy mẫu có hoàn lại (bootstrap sampling), tức là từ tập dữ liệu gốc, một số mẫu được chọn ngẫu nhiên với khả năng lặp lại. Mỗi cây quyết định trong rừng sẽ được xây dựng từ một tập con khác nhau này, giúp giảm thiểu tính đồng nhất và cải thiện khả năng tổng quát của mô hình.
* **Phân chia ngẫu nhiên**: Khi xây dựng mỗi cây quyết định, rừng ngẫu nhiên chỉ xem xét một tập con ngẫu nhiên của các thuộc tính trong quá trình phân chia. Điều này giúp giảm thiểu sự tương quan giữa các cây, dẫn đến một mô hình tổng thể mạnh mẽ hơn.
* **Kết hợp dự đoán**: Đối với bài toán phân loại, rừng ngẫu nhiên sử dụng phương pháp bỏ phiếu (voting) để xác định lớp cuối cùng dựa trên số lượng phiếu của các cây quyết định. Đối với bài toán hồi quy, giá trị dự đoán cuối cùng được tính là trung bình của các dự đoán từ các cây.

***2.3.2 Cấu trúc và quy trình xây dựng rừng ngẫu nhiên***

Quy trình xây dựng một mô hình rừng ngẫu nhiên bao gồm các bước sau:

1. **Thu thập và tiền xử lý dữ liệu**: Giống như bất kỳ mô hình máy học nào khác, bước đầu tiên là thu thập dữ liệu liên quan đến nhu cầu hàng tồn kho. Sau đó, dữ liệu cần được làm sạch và chuẩn bị, bao gồm xử lý giá trị thiếu, xử lý các giá trị ngoại lai và chuyển đổi các biến phân loại thành dạng số.
2. **Tạo mẫu dữ liệu**: Sử dụng kỹ thuật lấy mẫu bootstrap, tạo ra nhiều tập con dữ liệu từ tập dữ liệu gốc. Mỗi tập con này sẽ được sử dụng để xây dựng một cây quyết định riêng biệt trong rừng ngẫu nhiên.
3. **Xây dựng cây quyết định**: Với mỗi tập con dữ liệu, một cây quyết định được xây dựng. Trong quá trình xây dựng cây, chỉ một số thuộc tính được xem xét tại mỗi bước phân chia, giúp giảm thiểu sự tương quan giữa các cây.
4. **Kết hợp các cây**: Sau khi tất cả các cây quyết định đã được xây dựng, kết quả của chúng được kết hợp để đưa ra dự đoán cuối cùng. Đối với bài toán phân loại, dự đoán được thực hiện bằng cách bỏ phiếu, trong khi đối với bài toán hồi quy, giá trị được tính bằng trung bình của các giá trị dự đoán.
5. **Đánh giá mô hình**: Để đánh giá hiệu quả của rừng ngẫu nhiên, người ta thường sử dụng các chỉ số như độ chính xác (accuracy), độ nhạy (sensitivity), và độ đặc hiệu (specificity). Các kỹ thuật như phân chia tập huấn luyện và tập kiểm tra cũng được áp dụng để đảm bảo mô hình không bị quá khớp.

***2.3.3 Ưu điểm của rừng ngẫu nhiên***

Rừng ngẫu nhiên mang lại nhiều lợi ích đáng kể so với các mô hình khác, bao gồm:

* **Khả năng tổng quát cao**: Bằng cách kết hợp nhiều cây quyết định, rừng ngẫu nhiên có khả năng tổng quát tốt hơn so với một cây đơn lẻ. Điều này giúp giảm thiểu hiện tượng quá khớp và cải thiện độ chính xác của dự đoán.
* **Khả năng xử lý dữ liệu lớn và nhiều thuộc tính**: Rừng ngẫu nhiên có thể xử lý tốt các tập dữ liệu lớn với nhiều thuộc tính khác nhau. Nó có khả năng tự động phát hiện các mối quan hệ giữa các biến mà không cần phải xác định trước.
* **Chống lại hiện tượng quá khớp**: Rừng ngẫu nhiên có khả năng chống lại hiện tượng quá khớp tốt hơn nhờ vào việc kết hợp nhiều mô hình. Điều này giúp cải thiện khả năng tổng quát của mô hình và làm cho nó đáng tin cậy hơn trong các tình huống thực tế.
* **Đánh giá tầm quan trọng của thuộc tính**: Một trong những ưu điểm nổi bật của rừng ngẫu nhiên là khả năng đánh giá tầm quan trọng của các thuộc tính trong dữ liệu. Điều này giúp người phân tích nhận biết được các yếu tố ảnh hưởng lớn đến nhu cầu hàng tồn kho và có thể điều chỉnh các chiến lược kinh doanh tương ứng.
* **Không yêu cầu chuẩn hóa dữ liệu**: Rừng ngẫu nhiên không yêu cầu dữ liệu phải được chuẩn hóa, điều này giúp đơn giản hóa quy trình tiền xử lý.

***2.3.4 Hạn chế của rừng ngẫu nhiên***

Mặc dù có nhiều ưu điểm, rừng ngẫu nhiên cũng tồn tại một số hạn chế:

* **Tính không giải thích được**: Mặc dù cây quyết định đơn lẻ dễ hiểu và dễ giải thích, nhưng khi kết hợp nhiều cây trong rừng, việc hiểu rõ cách thức hoạt động của mô hình trở nên phức tạp hơn. Điều này có thể gây khó khăn trong việc truyền đạt kết quả cho các bên liên quan.
* **Thời gian tính toán lâu**: Việc xây dựng nhiều cây quyết định và kết hợp kết quả có thể đòi hỏi thời gian tính toán lớn hơn so với các mô hình đơn giản hơn, đặc biệt đối với các tập dữ liệu rất lớn.
* **Yêu cầu tài nguyên tính toán lớn**: Rừng ngẫu nhiên có thể yêu cầu nhiều tài nguyên tính toán, đặc biệt khi số lượng cây và số lượng thuộc tính tăng lên. Điều này có thể làm tăng chi phí tính toán và thời gian xử lý.

***2.3.5 Ứng dụng của rừng ngẫu nhiên trong dự báo hàng tồn kho***

Rừng ngẫu nhiên là một công cụ hữu ích trong việc dự đoán nhu cầu hàng tồn kho và có thể được áp dụng trong nhiều tình huống khác nhau:

* **Dự đoán nhu cầu hàng hóa**: Rừng ngẫu nhiên có khả năng phân tích nhiều yếu tố ảnh hưởng đến nhu cầu hàng hóa, từ đó đưa ra dự đoán chính xác về lượng hàng tồn kho cần thiết trong tương lai. Ví dụ, mô hình có thể sử dụng các yếu tố như mùa vụ, sự thay đổi trong hành vi tiêu dùng, và các yếu tố kinh tế để dự đoán nhu cầu.
* **Phân tích yếu tố ảnh hưởng**: Doanh nghiệp có thể sử dụng rừng ngẫu nhiên để xác định các yếu tố ảnh hưởng lớn đến nhu cầu hàng hóa, giúp tối ưu hóa quy trình sản xuất và phân phối.
* **Tối ưu hóa tồn kho**: Rừng ngẫu nhiên có thể được sử dụng để tối ưu hóa quy trình quản lý hàng tồn kho, giúp giảm thiểu chi phí và tăng cường hiệu quả hoạt động.
* **Phân loại sản phẩm**: Rừng ngẫu nhiên cũng có thể được sử dụng để phân loại các sản phẩm dựa trên các thuộc tính của chúng, từ đó giúp doanh nghiệp quyết định xem có nên tiếp tục sản xuất hay ngừng sản xuất một sản phẩm cụ thể.
* **Dự đoán xu hướng thị trường**: Với khả năng phân tích và phát hiện các mẫu trong dữ liệu, rừng ngẫu nhiên có thể giúp doanh nghiệp nhận biết các xu hướng thị trường và thay đổi trong nhu cầu hàng hóa, từ đó điều chỉnh chiến lược kinh doanh.

## 2.4 Học sâu (Deep Learning)

Học sâu (Deep Learning) là một nhánh của máy học, tập trung vào việc sử dụng các mạng nơ-ron (neural networks) với nhiều lớp để xử lý và phân tích dữ liệu phức tạp. Nhờ vào khả năng tự động học và phát hiện các đặc điểm từ dữ liệu mà không cần sự can thiệp của con người, học sâu đã trở thành một trong những công nghệ tiên tiến nhất trong lĩnh vực dự đoán, bao gồm cả dự báo hàng tồn kho.

***2.4.1 Đặc điểm của học sâu***

* **Mạng nơ-ron nhiều lớp (Deep Neural Networks)**:
  + Học sâu sử dụng mạng nơ-ron với nhiều lớp (hay còn gọi là mạng nơ-ron sâu). Các lớp này bao gồm lớp đầu vào (input layer), các lớp ẩn (hidden layers), và lớp đầu ra (output layer). Mỗi lớp có thể chứa hàng nghìn nơ-ron, cho phép mô hình phát hiện các mẫu phức tạp trong dữ liệu.
  + Các lớp ẩn giúp mô hình học các đặc trưng từ dữ liệu một cách tự động, từ những đặc trưng đơn giản ở các lớp gần đầu vào đến những đặc trưng phức tạp hơn ở các lớp gần đầu ra.
* **Khả năng xử lý dữ liệu phi cấu trúc**:
  + Một trong những ưu điểm nổi bật của học sâu là khả năng xử lý các dữ liệu phi cấu trúc như hình ảnh, âm thanh, và văn bản. Điều này mở rộng khả năng áp dụng của học sâu trong nhiều lĩnh vực khác nhau, bao gồm cả dự báo nhu cầu hàng tồn kho dựa trên hành vi mua sắm của khách hàng.
* **Khả năng học từ dữ liệu lớn**:
  + Học sâu có khả năng tự học từ một lượng lớn dữ liệu mà không cần phải tạo ra các đặc trưng thủ công. Điều này cho phép mô hình nắm bắt được các mối quan hệ phức tạp và tương tác giữa các biến.

***2.4.2 Cấu trúc của mạng nơ-ron***

* **Lớp đầu vào (Input Layer)**:
  + Lớp đầu vào nhận dữ liệu thô (raw data) từ bên ngoài. Mỗi nơ-ron trong lớp này đại diện cho một thuộc tính (feature) của dữ liệu.
* **Các lớp ẩn (Hidden Layers)**:
  + Các lớp ẩn nằm giữa lớp đầu vào và lớp đầu ra. Mỗi nơ-ron trong các lớp này thực hiện các phép toán và sử dụng hàm kích hoạt (activation function) để quyết định xem nơ-ron đó có "kích hoạt" hay không. Các hàm kích hoạt phổ biến bao gồm ReLU (Rectified Linear Unit), Sigmoid, và Tanh.
  + Số lượng lớp ẩn và số lượng nơ-ron trong mỗi lớp có thể thay đổi tùy thuộc vào độ phức tạp của bài toán và kích thước dữ liệu.
* **Lớp đầu ra (Output Layer)**:
  + Lớp đầu ra sản sinh ra kết quả cuối cùng của mô hình. Đối với bài toán phân loại, lớp đầu ra có thể sử dụng hàm kích hoạt Softmax để tạo ra xác suất cho từng lớp.

***2.4.3 Quá trình huấn luyện mô hình***

* **Dữ liệu huấn luyện**:
  + Quá trình huấn luyện mô hình học sâu yêu cầu một lượng lớn dữ liệu huấn luyện. Dữ liệu này cần được chuẩn bị kỹ lưỡng, bao gồm cả việc phân chia thành dữ liệu huấn luyện, dữ liệu xác thực (validation), và dữ liệu kiểm tra (test).
* **Thuật toán tối ưu hóa**:
  + Mô hình học sâu thường sử dụng các thuật toán tối ưu hóa như Stochastic Gradient Descent (SGD), Adam, hay RMSprop để cập nhật các trọng số (weights) của nơ-ron. Quá trình tối ưu hóa nhằm mục tiêu giảm thiểu hàm mất mát (loss function) – một thước đo về độ sai lệch giữa dự đoán của mô hình và giá trị thực tế.
* **Điều chỉnh tham số**:
  + Việc điều chỉnh các tham số như tỷ lệ học (learning rate), kích thước batch (batch size), và số epoch (số lần huấn luyện) là rất quan trọng để đảm bảo mô hình không bị overfitting hoặc underfitting.

***2.4.4 Ứng dụng của học sâu trong dự báo hàng tồn kho***

* **Dự đoán nhu cầu dựa trên hành vi khách hàng**:
  + Học sâu có thể được sử dụng để phân tích các dữ liệu phức tạp từ hành vi mua sắm của khách hàng. Ví dụ, mô hình có thể dự đoán nhu cầu hàng tồn kho dựa trên lịch sử mua sắm, thời gian trong năm, và các yếu tố khác như thời tiết hoặc sự kiện đặc biệt.
* **Phân tích hình ảnh và video**:
  + Trong một số ngành công nghiệp, học sâu có thể được áp dụng để phân tích hình ảnh hoặc video để theo dõi mức độ tồn kho và tình trạng hàng hóa. Ví dụ, các mô hình học sâu có thể nhận diện sản phẩm trên kệ hàng và cung cấp thông tin về mức tồn kho thực tế.
* **Tối ưu hóa chuỗi cung ứng**:
  + Học sâu cũng có thể được sử dụng để tối ưu hóa quy trình cung ứng, từ việc dự đoán nhu cầu cho đến lập kế hoạch sản xuất và phân phối. Bằng cách kết hợp dữ liệu từ nhiều nguồn khác nhau, mô hình có thể đưa ra những quyết định tối ưu hơn về việc đặt hàng và lưu kho.

***2.4.5 Thách thức trong việc áp dụng học sâu***

* **Cần nhiều dữ liệu**:
  + Học sâu yêu cầu một lượng lớn dữ liệu để có thể hoạt động hiệu quả. Trong một số trường hợp, việc thu thập dữ liệu đủ lớn có thể là một thách thức, đặc biệt là đối với các doanh nghiệp nhỏ.
* **Chi phí tính toán**:
  + Việc huấn luyện các mô hình học sâu thường đòi hỏi tài nguyên tính toán lớn, bao gồm cả GPU hoặc TPU. Điều này có thể tạo ra chi phí cao cho các doanh nghiệp không có hạ tầng công nghệ phù hợp.
* **Khó khăn trong việc giải thích**:
  + Một trong những vấn đề chính với học sâu là khả năng giải thích. Các mô hình học sâu thường được coi là "hộp đen", khó xác định lý do cụ thể cho các dự đoán mà chúng đưa ra. Điều này có thể gây khó khăn cho việc ra quyết định trong các lĩnh vực yêu cầu độ tin cậy cao.

## 2.5 Mô hình hồi quy logistic (Logistic Regression)

Hồi quy logistic là một mô hình thống kê phổ biến được sử dụng trong máy học để dự đoán xác suất mà một biến phụ thuộc thuộc về một trong hai lớp (biến phân loại nhị phân). Khác với hồi quy tuyến tính, hồi quy logistic không dự đoán giá trị thực mà thay vào đó, nó dự đoán xác suất thuộc về một lớp nhất định.

***2.5.1 Nguyên lý hoạt động***

Mô hình hồi quy logistic hoạt động dựa trên việc sử dụng một hàm logistic, hay còn gọi là hàm sigmoid, để chuyển đổi đầu ra của hồi quy tuyến tính thành xác suất. Hàm này có hình dạng chữ S, cho phép nó xác định được xác suất mà một quan sát thuộc về một lớp nhất định. Cụ thể, khi giá trị đầu vào lớn, xác suất dự đoán sẽ tiến gần đến 1, và khi giá trị đầu vào nhỏ, xác suất sẽ tiến gần đến 0.

Hồi quy logistic hoạt động bằng cách tìm mối quan hệ giữa các biến độc lập (các yếu tố ảnh hưởng) và biến phụ thuộc (nhu cầu hàng hóa). Mô hình này xác định cách thức mà mỗi yếu tố độc lập tác động đến xác suất mà một quan sát thuộc về một lớp cụ thể.

***2.5.2 Cách thức áp dụng trong dự báo hàng tồn kho***

Trong bối cảnh dự báo hàng tồn kho, hồi quy logistic có thể được sử dụng để xác định khả năng mà một sản phẩm sẽ được tiêu thụ trong một khoảng thời gian nhất định. Ví dụ, mô hình có thể được áp dụng để dự đoán xem một sản phẩm có khả năng bán hết hay không dựa trên các yếu tố như:

* **Khuyến mãi**: Mức độ giảm giá hoặc chương trình khuyến mãi có thể ảnh hưởng đến khả năng tiêu thụ của sản phẩm.
* **Thời gian**: Các yếu tố thời vụ như ngày lễ, sự kiện đặc biệt có thể làm tăng nhu cầu tiêu thụ.
* **Đặc điểm sản phẩm**: Các thuộc tính như kích thước, màu sắc, thương hiệu có thể ảnh hưởng đến sự lựa chọn của khách hàng.

Bằng cách sử dụng hồi quy logistic, doanh nghiệp có thể phân loại sản phẩm thành hai nhóm: sản phẩm có khả năng tiêu thụ cao và sản phẩm có khả năng tiêu thụ thấp. Điều này sẽ giúp doanh nghiệp đưa ra quyết định tốt hơn trong việc quản lý hàng tồn kho và tối ưu hóa quy trình sản xuất.

***2.5.3 Ưu điểm của hồi quy logistic***

* **Dễ hiểu và dễ triển khai**: Hồi quy logistic là một trong những mô hình đơn giản nhất trong máy học, dễ dàng giải thích và triển khai trong thực tế.
* **Xác suất dự đoán**: Mô hình cung cấp xác suất cho mỗi lớp, cho phép doanh nghiệp có cái nhìn tổng quát về tình hình tiêu thụ hàng hóa.
* **Khả năng xử lý dữ liệu phi tuyến**: Hồi quy logistic có thể xử lý các biến không tuân theo mối quan hệ tuyến tính, cho phép mô hình hóa các tình huống phức tạp hơn trong dự báo.

***2.5.4 Hạn chế của hồi quy logistic***

* **Giả định tuyến tính**: Hồi quy logistic giả định rằng có một mối quan hệ tuyến tính giữa các biến độc lập và logit của biến phụ thuộc. Nếu mối quan hệ không tuân theo giả định này, độ chính xác của mô hình có thể bị giảm.
* **Chỉ áp dụng cho hai lớp**: Hồi quy logistic chỉ áp dụng cho các bài toán phân loại nhị phân. Đối với các bài toán phân loại nhiều lớp, cần phải áp dụng các phương pháp khác như hồi quy logistic đa lớp.
* **Yêu cầu dữ liệu lớn**: Mô hình hồi quy logistic yêu cầu một lượng lớn dữ liệu để đảm bảo rằng các kết quả dự đoán là chính xác và đáng tin cậy.

## 2.6 Mô hình ARIMA (AutoRegressive Integrated Moving Average)

Mô hình ARIMA (AutoRegressive Integrated Moving Average) là một trong những mô hình phổ biến nhất trong phân tích chuỗi thời gian. Nó được sử dụng để dự đoán các biến theo thời gian, đặc biệt là trong các lĩnh vực như tài chính, kinh tế và quản lý chuỗi cung ứng. Mô hình này có khả năng nắm bắt các xu hướng và mùa vụ trong dữ liệu, giúp doanh nghiệp dự đoán nhu cầu hàng tồn kho một cách chính xác hơn.

***2.6.1 Cấu trúc của mô hình ARIMA***

Mô hình ARIMA được xây dựng dựa trên ba thành phần chính:

1. **AR (AutoRegressive)**: Phần này cho thấy rằng giá trị hiện tại của chuỗi thời gian phụ thuộc vào một hoặc nhiều giá trị trước đó. Ký hiệu p thể hiện số lượng lag (số lần trễ) được sử dụng trong mô hình.
2. **I (Integrated)**: Phần này đề cập đến việc lấy sai phân của chuỗi thời gian để làm cho nó ổn định. Ký hiệu d là số lần lấy sai phân cần thiết để làm cho chuỗi thời gian ổn định.
3. **MA (Moving Average)**: Phần này thể hiện rằng giá trị hiện tại của chuỗi thời gian phụ thuộc vào các sai số ngẫu nhiên của các giá trị trước đó. Ký hiệu q thể hiện số lượng lag của sai số trong mô hình.

Mô hình ARIMA thường được ký hiệu là ARIMA(p, d, q), trong đó:

* **p**: số lượng lag trong phần hồi quy tự hồi.
* **d**: số lần sai phân để làm cho chuỗi ổn định.
* **q**: số lượng lag của sai số trong phần trung bình động.

***2.6.2 Các bước xây dựng mô hình ARIMA***

Quá trình xây dựng mô hình ARIMA bao gồm các bước chính sau:

1. **Phân tích dữ liệu**:
   * Trước khi áp dụng mô hình ARIMA, cần phân tích và trực quan hóa dữ liệu để xác định các xu hướng và mùa vụ. Sử dụng đồ thị thời gian để nhận diện các đặc điểm này là bước quan trọng.
2. **Kiểm tra sự ổn định của chuỗi thời gian**:
   * Một chuỗi thời gian cần phải ổn định để áp dụng mô hình ARIMA. Các phương pháp kiểm tra sự ổn định bao gồm kiểm tra Dickey-Fuller (ADF test). Nếu chuỗi không ổn định, cần thực hiện lấy sai phân.
3. **Lấy sai phân**:
   * Sai phân là quá trình trừ giá trị hiện tại với giá trị trước đó để loại bỏ xu hướng. Số lần sai phân (d) cần thiết được xác định trong bước kiểm tra ổn định.
4. **Xác định các tham số p và q**:
   * Tham số p và q có thể được xác định bằng cách sử dụng các đồ thị ACF (Autocorrelation Function) và PACF (Partial Autocorrelation Function). Đồ thị ACF giúp xác định giá trị của q, trong khi PACF giúp xác định giá trị của p.
5. **Xây dựng mô hình**:
   * Sau khi xác định các tham số, mô hình ARIMA được xây dựng và huấn luyện trên dữ liệu.
6. **Đánh giá mô hình**:
   * Mô hình được đánh giá bằng cách sử dụng các chỉ số như AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), và kiểm tra độ chính xác trên tập dữ liệu kiểm tra.
7. **Dự đoán và kiểm tra**:
   * Sau khi mô hình được tối ưu hóa, có thể sử dụng để dự đoán nhu cầu hàng tồn kho. Kiểm tra dự đoán bằng cách so sánh với dữ liệu thực tế.

***2.6.3 Ứng dụng của mô hình ARIMA trong dự báo hàng tồn kho***

Mô hình ARIMA rất hữu ích trong việc dự đoán nhu cầu hàng tồn kho trong các tình huống sau:

* **Dữ liệu theo chu kỳ**: ARIMA có khả năng xử lý dữ liệu theo chu kỳ, giúp doanh nghiệp nhận diện và dự đoán các mùa vụ trong nhu cầu hàng hóa.
* **Dữ liệu không ổn định**: Mô hình ARIMA cho phép thực hiện sai phân để biến đổi dữ liệu không ổn định thành dữ liệu ổn định, giúp dự đoán chính xác hơn.
* **Dự đoán dài hạn**: ARIMA có thể được sử dụng để dự đoán nhu cầu hàng tồn kho trong dài hạn, điều này rất quan trọng trong việc lập kế hoạch sản xuất và lưu kho.

***2.6.4 Ưu điểm và nhược điểm của mô hình ARIMA***

* **Ưu điểm**:
  + Khả năng nắm bắt các yếu tố theo thời gian: ARIMA có thể mô hình hóa các xu hướng và mùa vụ, giúp dự đoán nhu cầu chính xác hơn.
  + Đơn giản trong việc triển khai: Quá trình xây dựng và áp dụng mô hình ARIMA tương đối đơn giản và dễ hiểu.
  + Hiệu quả với dữ liệu lịch sử: Mô hình này phát huy tốt hiệu quả khi có đủ dữ liệu lịch sử để huấn luyện.
* **Nhược điểm**:
  + Yêu cầu dữ liệu ổn định: Mô hình ARIMA yêu cầu chuỗi thời gian phải ổn định, điều này có thể gây khó khăn trong một số trường hợp.
  + Giới hạn trong việc mô hình hóa các yếu tố phi tuyến: ARIMA có thể không hoạt động tốt trong việc mô hình hóa các mối quan hệ phi tuyến hoặc phức tạp.
  + Phụ thuộc vào chất lượng dữ liệu: Độ chính xác của mô hình ARIMA phụ thuộc lớn vào chất lượng của dữ liệu đầu vào.

## 2.7 Học máy nâng cao (Ensemble Learning)

Học máy nâng cao (Ensemble Learning) là một kỹ thuật mạnh mẽ trong lĩnh vực máy học, cho phép kết hợp nhiều mô hình khác nhau để cải thiện độ chính xác và độ tin cậy của dự đoán. Thay vì chỉ sử dụng một mô hình duy nhất, phương pháp này kết hợp nhiều mô hình (còn gọi là "base learners") để đưa ra một dự đoán chung tốt hơn. Học máy nâng cao đã trở thành một trong những công cụ chính trong các ứng dụng thực tiễn, đặc biệt là trong dự báo nhu cầu hàng tồn kho.

***2.7.1 Cách thức hoạt động của Học máy nâng cao***

Học máy nâng cao hoạt động dựa trên nguyên tắc rằng sự kết hợp của nhiều mô hình khác nhau có thể tạo ra một mô hình tổng hợp mạnh mẽ hơn so với việc sử dụng bất kỳ mô hình đơn lẻ nào. Có hai kỹ thuật chính trong học máy nâng cao:

* **Bagging (Bootstrap Aggregating)**:
  + Bagging là một phương pháp mà trong đó nhiều mô hình được huấn luyện độc lập trên các tập dữ liệu con được chọn ngẫu nhiên từ tập dữ liệu gốc. Tập dữ liệu con được tạo ra bằng cách lấy mẫu với hoàn lại (bootstrap sampling), tức là các mẫu có thể được chọn nhiều lần.
  + Sau khi huấn luyện, dự đoán từ mỗi mô hình sẽ được kết hợp bằng cách tính trung bình (đối với các bài toán hồi quy) hoặc bỏ phiếu đa số (đối với các bài toán phân loại).
  + **Ví dụ**: Rừng ngẫu nhiên (Random Forest) là một ứng dụng nổi bật của Bagging, trong đó nhiều cây quyết định được tạo ra từ các mẫu ngẫu nhiên của dữ liệu.
* **Boosting**:
  + Boosting là một phương pháp trong đó các mô hình được huấn luyện tuần tự, mỗi mô hình mới sẽ cố gắng sửa chữa các sai sót của mô hình trước đó. Trong mỗi vòng lặp, trọng số của các mẫu dữ liệu được điều chỉnh để tăng cường ảnh hưởng của những mẫu mà mô hình trước đó đã dự đoán sai.
  + Sau khi hoàn thành quá trình huấn luyện, các dự đoán từ các mô hình sẽ được kết hợp để tạo ra một dự đoán cuối cùng, thường thông qua trung bình có trọng số.
  + **Ví dụ**: AdaBoost và Gradient Boosting là hai kỹ thuật Boosting phổ biến được áp dụng trong học máy.

***2.7.2 Ưu điểm của Học máy nâng cao***

* **Độ chính xác cao**: Kết hợp nhiều mô hình cho phép tận dụng sức mạnh của từng mô hình, từ đó cải thiện độ chính xác của dự đoán. Học máy nâng cao thường có hiệu suất vượt trội so với các mô hình đơn lẻ.
* **Khả năng chống overfitting**: Bằng cách kết hợp nhiều mô hình, học máy nâng cao có thể giảm thiểu hiện tượng overfitting, một vấn đề phổ biến trong việc xây dựng mô hình học máy, khi mô hình quá phức tạp và chỉ dự đoán tốt trên dữ liệu huấn luyện mà kém trên dữ liệu kiểm tra.
* **Khả năng xử lý dữ liệu không đồng nhất**: Học máy nâng cao có khả năng xử lý tốt các loại dữ liệu khác nhau và có thể được áp dụng cho nhiều bài toán khác nhau.
* **Dễ dàng giải thích**: Một số kỹ thuật học máy nâng cao như Random Forest có thể cung cấp thông tin về tầm quan trọng của các biến, giúp doanh nghiệp hiểu rõ hơn về các yếu tố ảnh hưởng đến nhu cầu hàng hóa.

***2.7.3 Ứng dụng của Học máy nâng cao trong Dự báo hàng tồn kho***

Học máy nâng cao đã được áp dụng rộng rãi trong nhiều lĩnh vực khác nhau, trong đó có dự báo hàng tồn kho. Cụ thể, các ứng dụng bao gồm:

* **Dự báo nhu cầu**: Sử dụng các mô hình học máy nâng cao để dự đoán nhu cầu hàng hóa dựa trên các yếu tố như lịch sử bán hàng, mùa vụ, và các chương trình khuyến mãi. Điều này giúp doanh nghiệp có kế hoạch sản xuất và lưu kho hợp lý.
* **Quản lý tồn kho**: Các mô hình học máy nâng cao có thể giúp doanh nghiệp tối ưu hóa quy trình quản lý tồn kho, xác định thời điểm đặt hàng và lượng hàng cần thiết để đáp ứng nhu cầu mà không gặp phải tình trạng thiếu hụt hoặc dư thừa.
* **Phân tích hành vi khách hàng**: Học máy nâng cao cũng có thể được sử dụng để phân tích hành vi mua sắm của khách hàng, từ đó dự đoán nhu cầu trong tương lai và điều chỉnh chiến lược tiếp thị một cách hiệu quả.
* **Phát hiện gian lận**: Trong một số trường hợp, các mô hình học máy nâng cao cũng có thể được áp dụng để phát hiện gian lận trong các giao dịch bán hàng, bảo vệ doanh nghiệp khỏi các rủi ro tài chính.

## 2.8 Mô hình phân tích chuỗi thời gian (Time Series Analysis)

Phân tích chuỗi thời gian là một kỹ thuật quan trọng trong lĩnh vực dự báo, chủ yếu được sử dụng để phân tích dữ liệu theo thời gian nhằm nhận diện các xu hướng, mùa vụ, và các yếu tố ngẫu nhiên. Trong dự báo hàng tồn kho, phân tích chuỗi thời gian giúp doanh nghiệp nắm bắt được các biến động trong nhu cầu theo thời gian, từ đó đưa ra các quyết định sản xuất và lưu kho hợp lý.

***2.8.1 Khái niệm cơ bản về chuỗi thời gian***

Chuỗi thời gian là một tập hợp các điểm dữ liệu được thu thập hoặc ghi lại theo thời gian. Dữ liệu chuỗi thời gian thường bao gồm các yếu tố như:

* **Giá trị**: Giá trị thực tế của biến được quan tâm (ví dụ: lượng hàng bán, số lượng hàng tồn kho).
* **Thời gian**: Khoảng thời gian mà dữ liệu được thu thập (ví dụ: ngày, tuần, tháng, năm).

***2.8.2 Các thành phần của chuỗi thời gian***

Chuỗi thời gian thường được phân tích qua các thành phần chính sau:

* **Xu hướng (Trend)**: Xu hướng thể hiện sự biến đổi của dữ liệu theo thời gian, có thể tăng, giảm hoặc không thay đổi. Việc nhận diện xu hướng là rất quan trọng trong dự báo, giúp doanh nghiệp hiểu được hướng đi của nhu cầu trong tương lai.
* **Mùa vụ (Seasonality)**: Mùa vụ là các biến động lặp đi lặp lại theo chu kỳ trong một khoảng thời gian nhất định. Ví dụ, nhu cầu hàng hóa có thể tăng vào mùa lễ hội hoặc vào các dịp lễ tết. Nhận diện mùa vụ giúp doanh nghiệp lên kế hoạch sản xuất và lưu kho theo chu kỳ.
* **Nhiễu (Noise)**: Nhiễu là các biến động ngẫu nhiên không thể dự đoán, có thể do nhiều yếu tố khác nhau gây ra, như sự thay đổi bất ngờ trong thị trường hoặc các yếu tố bên ngoài không lường trước được.
* **Chu kỳ (Cyclicality)**: Các biến động theo chu kỳ là những thay đổi diễn ra theo chu kỳ nhưng không theo một lịch trình cố định. Chúng thường liên quan đến các yếu tố kinh tế lớn như suy thoái hoặc tăng trưởng.

***2.8.3 Các phương pháp phân tích chuỗi thời gian***

Có nhiều phương pháp để phân tích chuỗi thời gian, dưới đây là một số phương pháp phổ biến:

* **Phân tích Decomposition**:
  + Phương pháp này phân chia chuỗi thời gian thành các thành phần cơ bản (xu hướng, mùa vụ, nhiễu) để hiểu rõ hơn về cấu trúc dữ liệu.
  + Cách tiếp cận này cho phép nhận diện và phân tách các yếu tố ảnh hưởng đến chuỗi thời gian, giúp doanh nghiệp điều chỉnh kế hoạch dự trữ hàng hóa.
* **Mô hình hồi quy chuỗi thời gian (Time Series Regression)**:
  + Hồi quy chuỗi thời gian sử dụng hồi quy để mô hình hóa và dự đoán giá trị của chuỗi thời gian dựa trên các yếu tố giải thích (independent variables).
  + Mô hình này có thể kết hợp nhiều yếu tố ảnh hưởng, như yếu tố mùa vụ, để cải thiện độ chính xác của dự báo.
* **Hàm trượt trung bình (Moving Average)**:
  + Hàm trượt trung bình giúp làm mịn dữ liệu chuỗi thời gian bằng cách tính trung bình của một số lượng điểm dữ liệu lân cận. Điều này giúp giảm thiểu các biến động ngẫu nhiên và xác định xu hướng chung của dữ liệu.
  + Các loại hàm trượt trung bình bao gồm trượt trung bình đơn giản (Simple Moving Average) và trượt trung bình trọng số (Weighted Moving Average).
* **Mô hình ARIMA (AutoRegressive Integrated Moving Average)**:
  + ARIMA là một trong những mô hình phổ biến nhất trong phân tích chuỗi thời gian. Mô hình này kết hợp các yếu tố tự hồi quy (autoregressive), tích phân (integrated), và trung bình động (moving average) để mô hình hóa chuỗi thời gian.
  + Mô hình ARIMA cần phải được điều chỉnh sao cho dữ liệu là ổn định, và nó rất hữu ích trong việc dự đoán nhu cầu hàng tồn kho theo thời gian.
* **Mô hình SARIMA (Seasonal ARIMA)**:
  + Mô hình SARIMA là một phiên bản mở rộng của ARIMA, cho phép xử lý các thành phần mùa vụ trong dữ liệu chuỗi thời gian. Điều này rất quan trọng trong dự báo hàng tồn kho khi có sự biến động theo mùa trong nhu cầu.
* **Hàm hồi quy Exponential Smoothing**:
  + Đây là phương pháp dự đoán dựa trên các trọng số giảm dần theo thời gian, giúp tạo ra các dự đoán ngắn hạn rất hiệu quả. Nó rất linh hoạt và có thể áp dụng cho dữ liệu có xu hướng hoặc có mùa vụ.

***2.8.4 Ứng dụng của phân tích chuỗi thời gian trong dự báo hàng tồn kho***

Phân tích chuỗi thời gian đóng vai trò quan trọng trong việc dự báo nhu cầu hàng tồn kho. Việc sử dụng các mô hình phân tích chuỗi thời gian giúp doanh nghiệp:

* **Dự đoán chính xác hơn**: Giúp doanh nghiệp có thể dự đoán chính xác hơn về nhu cầu hàng hóa trong tương lai, từ đó cải thiện kế hoạch sản xuất và lưu kho.
* **Lập kế hoạch sản xuất**: Nhận diện các xu hướng và mùa vụ trong nhu cầu giúp doanh nghiệp điều chỉnh sản xuất và phân phối hợp lý.
* **Tối ưu hóa tồn kho**: Giảm thiểu tình trạng thừa hoặc thiếu hàng hóa, từ đó giảm chi phí lưu kho và tăng cường hiệu quả hoạt động.
* **Đưa ra quyết định nhanh chóng**: Cung cấp thông tin thời gian thực giúp doanh nghiệp nhanh chóng đưa ra quyết định phù hợp với biến động của thị trường.

**KẾT LUẬN CHƯƠNG 2**

Chương này đã trình bày một số mô hình máy học phổ biến trong dự báo hàng tồn kho, từ hồi quy đơn giản đến các kỹ thuật học sâu phức tạp. Mỗi mô hình có những ưu điểm và ứng dụng riêng, và việc lựa chọn mô hình phù hợp là rất quan trọng để tối ưu hóa quá trình dự báo.

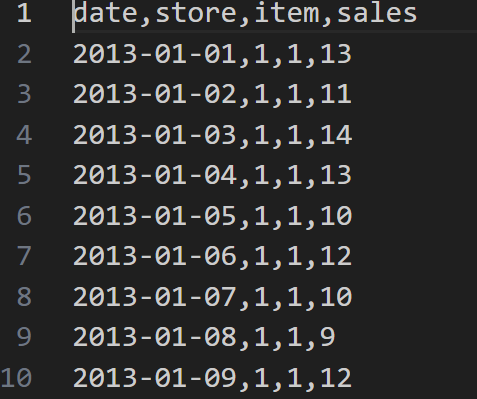
Bằng việc áp dụng các mô hình máy học trong dự báo hàng tồn kho, các doanh nghiệp có thể nâng cao độ chính xác của dự đoán, tối ưu hóa quy trình và giảm thiểu chi phí, từ đó cải thiện hiệu quả hoạt động và tăng cường khả năng cạnh tranh trên thị trường. Trong chương tiếp theo, chúng ta sẽ đi sâu vào phân tích chương trình cụ thể mà chúng ta sẽ áp dụng để dự báo nhu cầu hàng tồn kho bằng cách sử dụng máy học.

# **CHƯƠNG 3: PHÂN TÍCH CHƯƠNG TRÌNH DỰ BÁO NHU CẦU HÀNG TỒN KHO BẰNG CÁCH SỬ DỤNG MÁY HỌC**

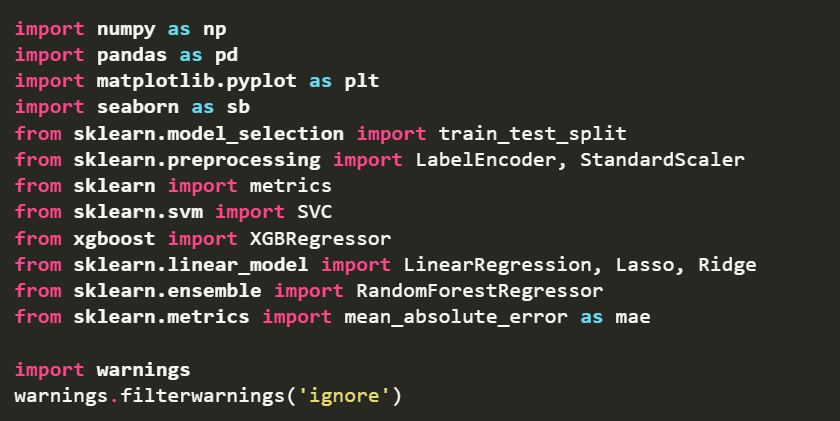
## 3.1. Mô tả chi tiết chương trình

- Trong chương trình, chúng ta sẽ thử triển khai một mô hình học máy có thể dự đoán số lượng hàng tồn kho cho các sản phẩm khác nhau được bán tại các cửa hàng khác nhau.

- Đầu tiên chúng tôi thu thập tập tin dữ liệu và đặt tên là **train.csv**. dữ liệu thu thập được có dạng như sau:



Sau khi thu thập được dữ liệu, ta sử dụng python để khởi tạo chương trình dự báo nhu cầu hàng tồn kho.

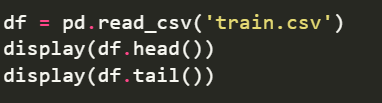


Các thư viện Python giúp chúng ta dễ dàng xử lý dữ liệu và thực hiện các tác vụ thông thường cũng như phức tạp chỉ với một dòng mã.

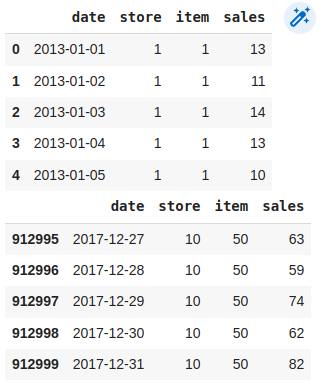
* **Pandas** – Thư viện này giúp tải khung dữ liệu ở dạng mảng 2D và có nhiều hàm để thực hiện các tác vụ phân tích một cách nhanh chóng.
* **Numpy** – Các mảng Numpy rất nhanh và có thể thực hiện các tính toán lớn trong thời gian rất ngắn.
* **Matplotlib/Seaborn** – Thư viện này được sử dụng để vẽ các hình minh họa trực quan.
* **Sklearn** – Mô-đun này chứa nhiều thư viện với các hàm đã được triển khai sẵn để thực hiện các tác vụ từ tiền xử lý dữ liệu đến phát triển và đánh giá mô hình.
* **XGBoost** – Thư viện này chứa thuật toán học máy eXtreme Gradient Boosting, là một trong những thuật toán giúp chúng ta đạt độ chính xác cao trong dự đoán.

- Bây giờ chúng ta hãy tải tập dữ liệu vào khung dữ liệu của pandas và in ra năm hàng đầu tiên của nó.

- Dưới đây là đoạn mã Python để thực hiện điều này:



-Ta sẽ được kết quả như sau:



Như chúng ta có thể thấy, chúng ta có dữ liệu trong năm năm cho 10 cửa hàng và 50 sản phẩm, tính toán như sau:

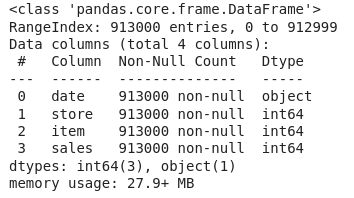
**(365 \* 4 + 366) \* 10 \* 50 = 913000**

ta sử dụng lệnh: **df.shape** ta thu được kết quả: **(913000, 4)**



Hãy kiểm tra xem cột nào trong tập dữ liệu chứa loại dữ liệu nào bằng cách sử dụng lệnh **df.info()**

Ta được kết quả đầu ra như sau:



Dựa trên thông tin ở trên về dữ liệu trong mỗi cột, chúng ta có thể nhận thấy rằng không có giá trị null.

Sau đó ta sẽ sử dụng lệnh **df.describe()** sử dụng để tạo ra các thống kê tóm tắt cho các cột dữ liệu số trong khung dữ liệu (**DataFrame**).

A screenshot of a white background with numbers

Description automatically generated

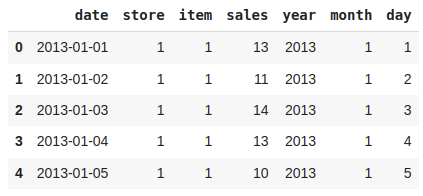
**Feature Engineering**

Trong một số trường hợp, nhiều đặc trưng có thể được tích hợp trong cùng một cột dữ liệu hoặc chúng ta cần tạo ra các đặc trưng mới từ những dữ liệu hiện có. Chúng ta sẽ cố gắng thêm một số đặc trưng bổ sung vào tập dữ liệu để có thể rút ra những thông tin hữu ích từ dữ liệu về hàng tồn kho. Nếu những đặc trưng này có ý nghĩa, chúng sẽ đóng vai trò quan trọng trong việc cải thiện đáng kể độ chính xác của mô hình dự báo.

A screen shot of a computer code

Description automatically generated

Đoạn mã này hữu ích trong việc trích xuất thông tin từ cột "date", cho phép phân tích và xử lý dữ liệu theo năm, tháng và ngày, từ đó có thể sử dụng thông tin này cho các mô hình dự báo hàng tồn kho trong dự án. Ta được đầu ra như sau:

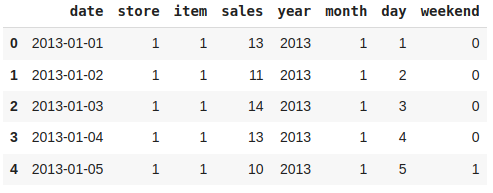


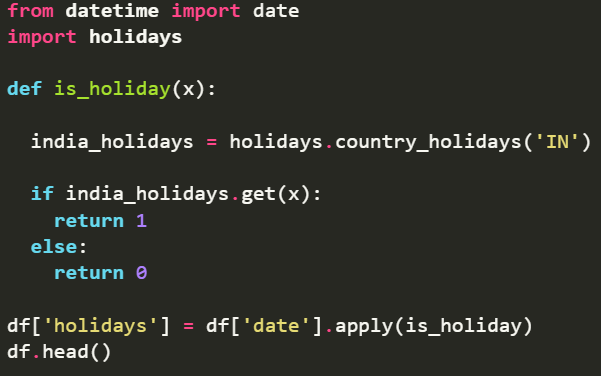
Có thể thấy rằng việc đó là ngày cuối tuần hay ngày trong tuần chắc chắn sẽ ảnh hưởng đến nhu cầu để đáp ứng các yêu cầu.

Đoạn tiếp theo để phân tích ảnh hưởng của các ngày cuối tuần và ngày trong tuần đến nhu cầu hàng tồn kho, giúp ta rút ra những thông tin quan trọng cho mô hình dự báo



Ta thu được kết quả:



Để tối ưu hơn ta sẽ làm thêm một cột có thể chỉ ra liệu có ngày lễ nào vào một ngày cụ thể

Ta sẽ thu được kết quả:



Bây giờ, chúng ta sẽ bổ sung một số đặc trưng tuần hoàn vào mô hình để cải thiện khả năng dự đoán nhu cầu hàng tồn kho

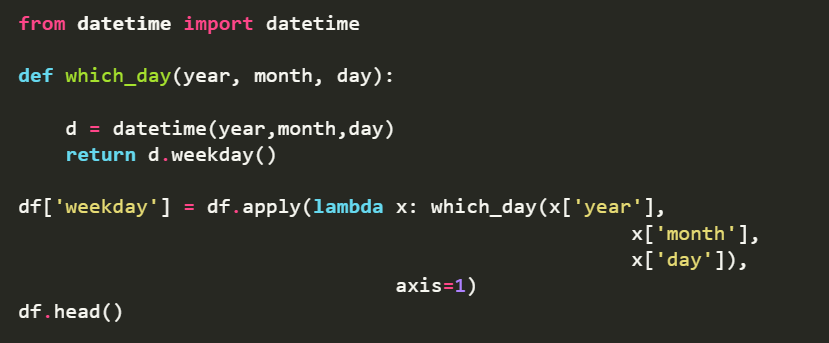
A black screen with yellow text

Description automatically generated

ta sẽ có kết quả:



Hãy thêm một cột vào mô hình để chỉ ra ngày trong tuần, từ đó giúp phân tích và dự đoán một cách chính xác hơn.



ta sẽ có kết quả như sau:



Ta sẽ tối ưu chương trình bằng lệnh:

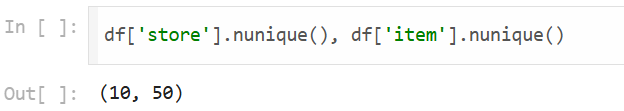


Có thể có một số đặc trưng liên quan khác cũng có thể được thêm vào tập dữ liệu này, nhưng trước mắt, chúng ta hãy cố gắng xây dựng mô hình với những đặc trưng hiện tại và rút ra một số thông tin hữu ích

**Exploratory Data Analysis**

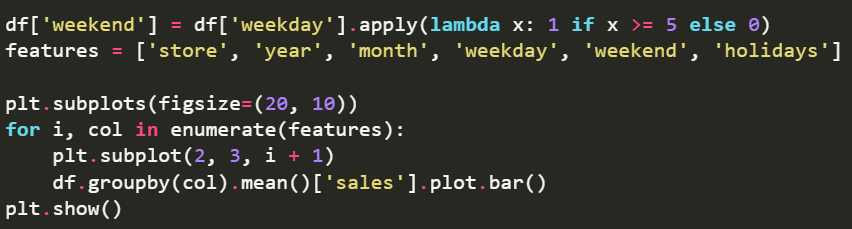
EDA là một phương pháp phân tích dữ liệu thông qua các kỹ thuật trực quan. Nó được sử dụng để phát hiện các xu hướng, mẫu hoặc kiểm tra giả định với sự hỗ trợ của các tóm tắt thống kê và các biểu diễn đồ họa.

Chúng ta đã thêm một số đặc trưng vào tập dữ liệu dựa trên các giả định. Bây giờ, hãy cùng kiểm tra mối quan hệ giữa các đặc trưng khác nhau với đặc trưng mục tiêu để tìm hiểu sâu hơn về dữ liệu và dự đoán nhu cầu hàng tồn kho.

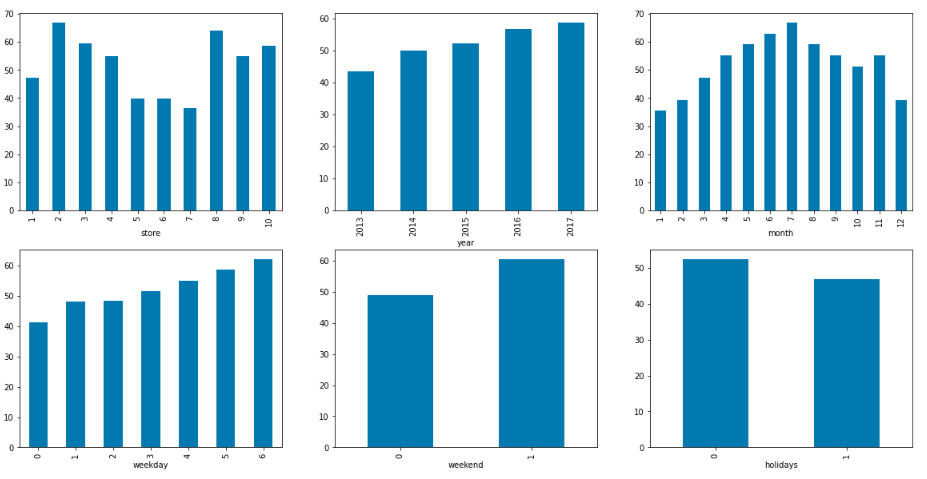
****

Từ đây, chúng ta có thể kết luận rằng có 10 cửa hàng độc nhất và họ bán 50 sản phẩm khác nhau.

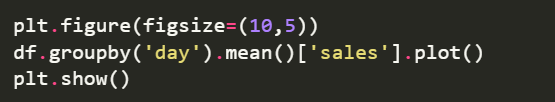
Tiếp theo chúng ta sử dụng đoạn mã sau để tạo ra biểu đồ cột hiển thị doanh số trung bình (sales) của từng đặc trưng trong danh sách features



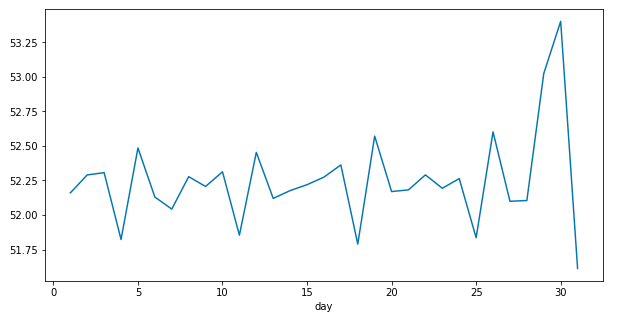
Ta sẽ được kết quả:



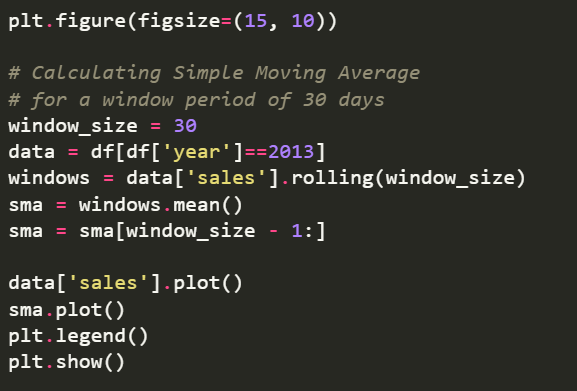
Bây giờ, hãy kiểm tra sự biến động của hàng tồn kho khi tháng gần đến ngày kết thúc

****

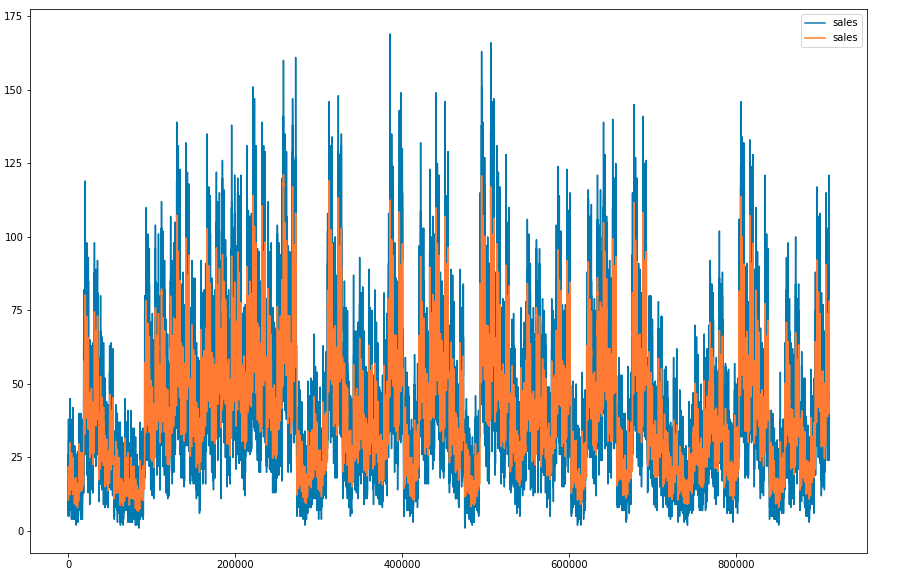
Ta thu được kết quả:

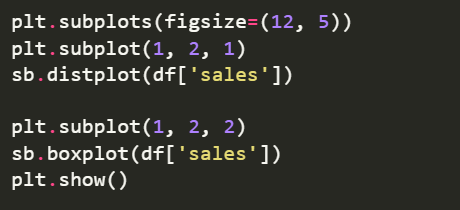


Ta sẽ vẽ đường trung bình động đơn giản (SMA) cho khoảng thời gian 30 ngày như sau:

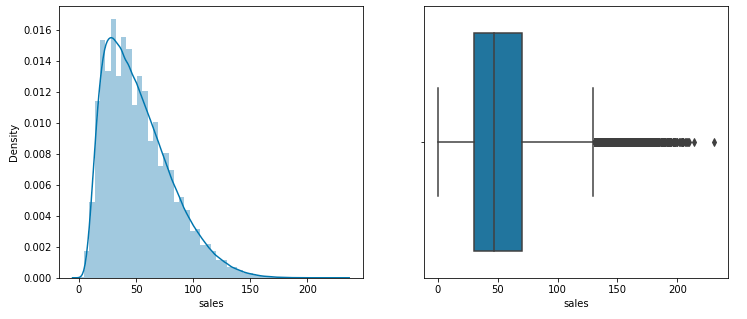


Ta sẽ được kết quả:



Vì dữ liệu trong cột doanh số là liên tục, hãy kiểm tra phân phối của nó để xác định xem có tồn tại các giá trị ngoại lai trong cột này hay không

Ta sẽ có kết quả:

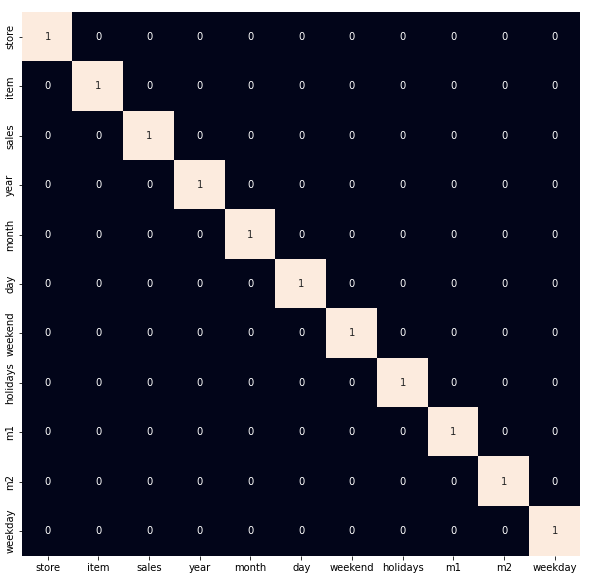


Các đặc trưng có mối tương quan cao có thể ảnh hưởng đáng kể đến độ chính xác của mô hình. Khi hai hoặc nhiều đặc trưng có mối liên hệ chặt chẽ với nhau, điều này có thể dẫn đến hiện tượng đa cộng tuyến, làm giảm khả năng giải thích của mô hình và gây khó khăn trong việc xác định ảnh hưởng của từng đặc trưng đối với biến mục tiêu. Do đó, việc phát hiện và xử lý các đặc trưng tương quan cao là rất quan trọng trong quá trình xây dựng mô hình dự báo hàng tồn kho.

A computer screen with white text

Description automatically generated

Ta có kết quả:

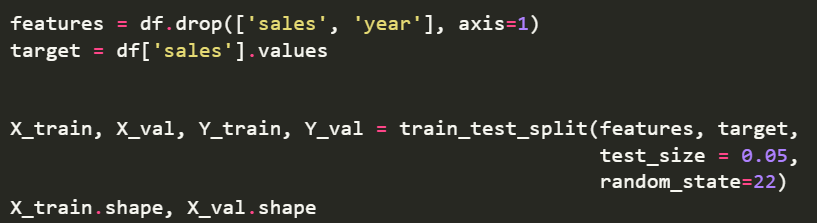


Như chúng ta đã quan sát trước đó, hãy loại bỏ các giá trị ngoại lai có mặt trong dữ liệu

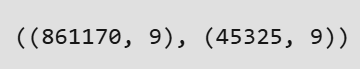


**Model Training**

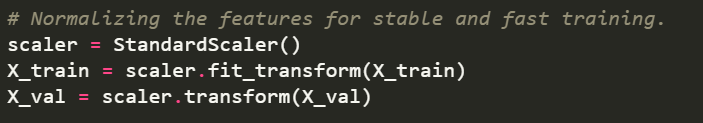
Bây giờ, chúng ta sẽ tách các đặc trưng và biến mục tiêu, sau đó chia chúng thành dữ liệu huấn luyện và dữ liệu kiểm tra. Sử dụng những dữ liệu này, chúng ta sẽ chọn mô hình hoạt động tốt nhất trên dữ liệu kiểm định.

****

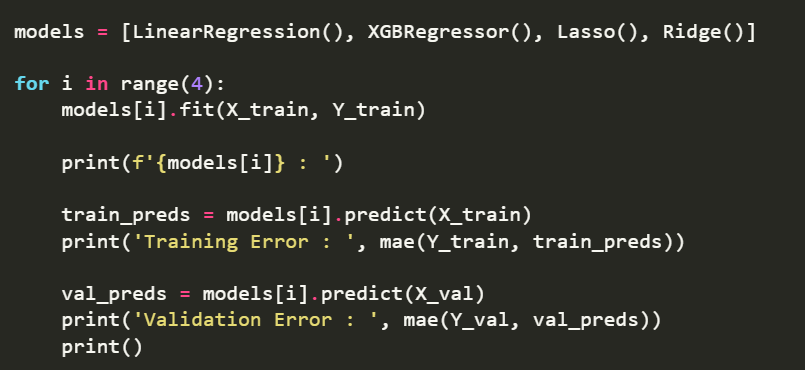
Ta được kết quả:

****

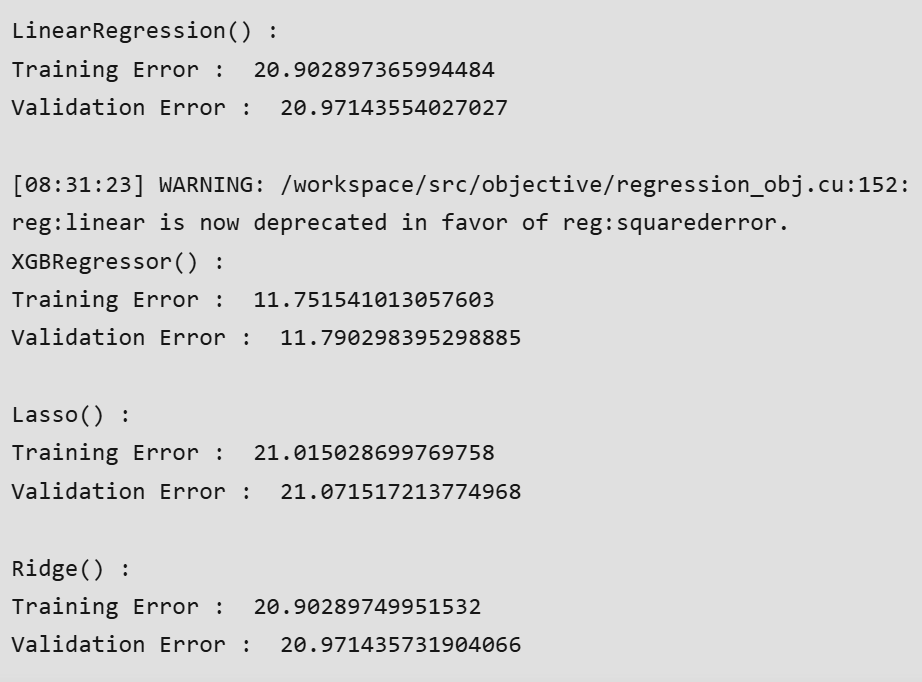
Việc chuẩn hóa dữ liệu trước khi đưa vào các mô hình machine learning giúp chúng ta đạt được quá trình huấn luyện ổn định và nhanh chóng

****

Chúng ta đã chia dữ liệu thành tập huấn luyện và tập kiểm định, đồng thời đã thực hiện việc chuẩn hóa dữ liệu. Bây giờ, hãy huấn luyện một số mô hình machine learning tiên tiến và chọn ra mô hình tốt nhất từ chúng bằng cách sử dụng tập dữ liệu kiểm định.

****

Ta sẽ thu được kết quả như sau:

****

# **TỔNG KẾT CHƯƠNG TRÌNH DỰ BÁO NHU CẦU HÀNG TỒN KHO**

Trong dự án này, chúng ta đã thực hiện một quy trình hoàn chỉnh để dự đoán nhu cầu hàng tồn kho bằng cách sử dụng các kỹ thuật machine learning. Đầu tiên, chúng ta đã nhập các thư viện cần thiết và tải dữ liệu vào khung dữ liệu pandas. Sau khi kiểm tra kích thước và cấu trúc của dữ liệu, chúng ta đã thực hiện phân tích khám phá dữ liệu (EDA) để xác định các đặc trưng quan trọng và loại bỏ các giá trị ngoại lai.

Tiếp theo, chúng ta đã tiến hành kỹ thuật kỹ thuật đặc trưng (feature engineering) bằng cách thêm các cột mới, như ngày tháng, ngày trong tuần, và các thông tin về ngày lễ. Chúng ta cũng đã tạo ra các đặc trưng tuần hoàn để phản ánh ảnh hưởng của thời gian lên nhu cầu hàng hóa.

Sau khi chuẩn bị dữ liệu, chúng ta đã tách dữ liệu thành các tập huấn luyện và kiểm định, thực hiện chuẩn hóa dữ liệu để đảm bảo quá trình huấn luyện ổn định và hiệu quả. Cuối cùng, chúng ta đã sẵn sàng huấn luyện và đánh giá một số mô hình machine learning tiên tiến, nhằm tìm ra mô hình tốt nhất cho dự đoán nhu cầu hàng tồn kho.

Thông qua quy trình này, chúng ta có thể khai thác dữ liệu hiệu quả hơn để đưa ra các quyết định dựa trên dự đoán chính xác về nhu cầu hàng tồn kho, từ đó giúp tối ưu hóa việc quản lý hàng hóa và nâng cao hiệu quả kinh doanh.

**TÀI LIỆU THAM KHẢO**

<https://www.geeksforgeeks.org/inventory-demand-forecasting-using-machine-learning-python/>

<https://www.kaggle.com/datasets/subho117/inventory-demand-forecasting-using-ml>

<https://media.geeksforgeeks.org/wp-content/uploads/20240909120859/Inventory_Demand_Forecasting_using_Machine_Learning_Python.ipynb>