**BỘ GIÁO DỤC VÀ ĐÀO TẠO**



**TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH**

**KHOA CÔNG NGHỆ THÔNG TIN**





**BÁO CÁO THỰC TẬP TỐT NGHIỆP**

**PHÂN TÍCH DỮ LIỆU KHÁCH HÀNG TRONG SẢN XUẤT NHUỘM**

**GIẢNG VIÊN HƯỚNG DẪN :   THS. VƯƠNG XUÂN CHÍ**

**SINH VIÊN THỰC HIỆN :   VÕ THÁI HIỂN**

**MSSV :   2000005690**

**LỚP :   20DTH1D**

**CHUYÊN NGÀNH :   KHOA HỌC DỮ LIỆU**

**KHÓA :   2020**

**ĐƠN VỊ THỰC TẬP :   CÔNG TY CỔ PHẦN PHẦN MỀM & CÔNG NGHỆ TỰ ĐỘNG 4.0 BẢO TÍN**

**Tp.HCM, Tháng  11  Năm  2023**

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**



**TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH**

**KHOA CÔNG NGHỆ THÔNG TIN**





**BÁO CÁO THỰC TẬP TỐT NGHIỆP**

**PHÂN TÍCH DỮ LIỆU KHÁCH HÀNG TRONG SẢN XUẤT NHUỘM**

**GIẢNG VIÊN HƯỚNG DẪN :   THS. VƯƠNG XUÂN CHÍ**

**SINH VIÊN THỰC HIỆN :   VÕ THÁI HIỂN**

**MSSV :   2000005690**

**LỚP :   20DTH1D**

**CHUYÊN NGÀNH :   KHOA HỌC DỮ LIỆU**

**KHÓA :   2020**

**ĐƠN VỊ THỰC TẬP :   CÔNG TY CỔ PHẦN PHẦN MỀM & CÔNG NGHỆ TỰ ĐỘNG 4.0 BẢO TÍN**

**Tp.HCM, Tháng  11  Năm  2023**

LỜI CẢM ƠN

Tôi xin gửi lời cảm ơn chân thành nhất đến toàn thể đội ngũ thành viên của Công Ty Cổ Phần Phần Mềm & Công Nghệ Tự Động 4.0 Bảo Tín, đặc biệt là anh Đặng Hồng Phú và thầy th.s Vương Xuân Chí, người đã làm cho chương trình thực tập của tôi trở thành một trải nghiệm học tập và nghề nghiệp không thể quên.

Anh Đặng Hồng Phú đã đóng vai trò quan trọng như một người hướng dẫn xuất sắc trong suốt thời gian thực tập của tôi tại Công Ty. Anh không chỉ chia sẻ sâu sắc kiến thức chuyên sâu về ngành công nghiệp mà còn hỗ trợ và tạo điều kiện để tôi áp dụng những kiến thức đó vào thực tế. Sự nhiệt huyết và tận tâm của anh đã tạo nên một môi trường làm việc tích cực và động lực cho sự phát triển cá nhân của tôi.

Ngoài ra, tôi cũng muốn bày tỏ lòng biết ơn sâu sắc đến thầy th.s Vương Xuân Chí, giảng viên hướng dẫn tận tâm. Thầy đã giúp đỡ tôi hiểu rõ hơn về lĩnh vực công nghệ và đưa ra những hướng dẫn chi tiết, giúp tôi vượt qua những thách thức trong quá trình học tập và làm việc.

Thực sự, thời gian tôi đã dành tại Công Ty Bảo Tín không chỉ là một kỳ thực tập mà còn là một giai đoạn quan trọng đối với sự phát triển cá nhân và sự hiểu biết về ngành công nghiệp. Tôi tin tưởng rằng những kinh nghiệm và kiến thức thu được từ đây sẽ là nền tảng vững chắc cho sự nghiệp tương lai của mình.

Cảm ơn các bạn đã thực chung nhóm đã đem đến những ý tưởng mới, sự sáng tạo và sự đầu tư tận tâm vào mọi dự án. Các bạn thực sự là những người đồng hành không thể thiếu để chúng ta có thể đạt được những thành công lớn.

Một lần nữa, tôi xin bày tỏ lòng biết ơn sâu sắc đến Công Ty Cổ Phần Phần Mềm & Công Nghệ Tự Động 4.0 Bảo Tín, anh Đặng Hồng Phú và thầy th.s Vương Xuân Chí. Hy vọng rằng tương lai sẽ mang lại nhiều thành công và phát triển hơn nữa cho công ty.

***Sinh viên thực hiện***

**Võ Thái Hiển**

NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN

1. Hình thức (Bố cục, trình bày, lỗi, các mục, hình, bảng, công thức, phụ lục, )

2. Nội dung (mục tiêu, phương pháp, kết quả, sao chép, các chương, tài liệu,..)

3. Kết luận

*TPHCM, Ngày …… tháng …… năm 2017*

**Giáo viên hướng dẫn**

(Ký tên, đóng dấu)

**Vương Xuân Chí**

NHẬN XÉT CỦA CƠ QUAN THỰC TẬP

*TPHCM, Ngày …… tháng …… năm 2017*

**Xác nhận của đơn vị thực tập**

(Ký tên, đóng dấu)

MỤC LỤC

Nội dung

[LỜI CẢM ƠN i](#_Toc154237825)

[NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN ii](#_Toc154237826)

[NHẬN XÉT CỦA CƠ QUAN THỰC TẬP iii](#_Toc154237827)

[MỤC LỤC iv](#_Toc154237828)

[DANH MỤC HÌNH vi](#_Toc154237829)

[DANH MỤC CÁC TỪ VIẾT TẮT vii](#_Toc154237830)

[CHƯƠNG 1./ GIỚI THIỆU CƠ QUAN THỰC TẬP VÀ ĐỀ TÀI 12](#_Toc154237831)

[1.1./ Giới thiệu cơ quan thực tập 12](#_Toc154237832)

[1.2./ Giới thiệu đề tài 15](#_Toc154237833)

[1.2.1./ Lý do chọn đề tài 15](#_Toc154237834)

[1.2.2./ Mục tiêu của đề tài 16](#_Toc154237835)

[1.2.3./ Đối tượng và phạm vi nguyên cứu của đề tài 16](#_Toc154237836)

[CHƯƠNG 2./ CƠ SỞ LÝ LUẬN VỀ VẤN ĐỀ NGHIÊN CỨU VÀ MÔ HÌNH LÝ THUYẾT 18](#_Toc154237837)

[2.1./ Cơ sở lý luận về vấn đề nghiên cứu 18](#_Toc154237838)

[2.2./ Cơ sở lý thuyết 20](#_Toc154237839)

[2.2.1/ Thu thập dữ liệu 20](#_Toc154237840)

[2.2.2./ Tiền xử lý dữ liệu 21](#_Toc154237841)

[2.2.3./ Phân tích dữ liệu 22](#_Toc154237842)

[2.2.4./ Tổng quan về Regression trong Machine Learning 23](#_Toc154237843)

[CHƯƠNG 3./ MÔ HÌNH THỰC NGHIỆM 26](#_Toc154237844)

[3.1./ Bộ dữ liệu 26](#_Toc154237845)

[3.1.1/ Bộ dữ liệu Chemical Using 26](#_Toc154237846)

[3.1.2./ Bộ dữ liệu Sale 26](#_Toc154237847)

[3.2./ Mô hình thực nghiệm 27](#_Toc154237848)

[3.2.1./ Bộ dữ liệu Chemical Using 28](#_Toc154237849)

[3.2.2./ Bộ dữ liệu Sale 36](#_Toc154237850)

[3.3./ Kết luận chung 40](#_Toc154237851)

[3.4./ Ưu điểm và nhược điểm 40](#_Toc154237852)

[3.5./ Phương hướng phát triển 41](#_Toc154237853)

[CHƯƠNG 4./ KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 43](#_Toc154237854)

[4.1./ Về bản thân 43](#_Toc154237855)

[4.2./ Về đề tài 43](#_Toc154237856)

[4.3./ Hướng phát triển cho bản thân 44](#_Toc154237857)

[TÀI LIỆU THAM KHẢO 46](#_Toc154237858)

DANH MỤC HÌNH

[Hình 1. 1: Hình ảnh của công ty giới thiệu 8](#_Toc154235679)

[Hình 3. 1: Ví dụ code thao tác cơ bản trên bộ dữ liệu 24](#_Toc154235684)

[Hình 3. 2: Kết quả của thao tác cơ bản 25](#_Toc154235685)

[Hình 3. 3: Code tạo biểu đồ so sánh 25](#_Toc154235686)

[Hình 3. 4: Biểu đồ so sánh 27](#_Toc154235687)

[Hình 3. 5: Thao tác đầu tiên cho mô hình dự đoán 27](#_Toc154235688)

[Hình 3. 6: Kết quả cột datatime mới 28](#_Toc154235689)

[Hình 3. 7: Xoá các cột không cần thiết 28](#_Toc154235690)

[Hình 3. 8: Kết quả sau khi xoá 28](#_Toc154235691)

[Hình 3. 9: Chia dữ liệu cho cột x,y 29](#_Toc154235692)

[Hình 3. 10: Code huấn luyện mô hình 30](#_Toc154235693)

[Hình 3. 11: Kiểm tra kết quả của mô hình 30](#_Toc154235694)

[Hình 3. 12: Kết quả kiểm tra mô hình 31](#_Toc154235695)

[Hình 3. 13: Thông kê id sản phẩm bán ra 32](#_Toc154235696)

[Hình 3. 14: Code thống kê sản phẩm bán chạy nhất và thấp nhất 32](#_Toc154235697)

[Hình 3. 15: Kết quả sản phẩm bán chạy nhất và thấp nhất 33](#_Toc154235698)

[Hình 3. 16: Lọc khách hàng 33](#_Toc154235699)

[Hình 3. 17: Chia cột thành ngày và thời gian 34](#_Toc154235700)

[Hình 3. 18: Kết quả sau khi chia 34](#_Toc154235701)

[Hình 3. 19: Code vẽ biểu đồ tổng quan cột TotalQuantity 35](#_Toc154235702)

[Hình 3. 20: Biểu đồ tổng quan cột TotalQuantity 35](#_Toc154235703)

DANH MỤC CÁC TỪ VIẾT TẮT

|  |  |  |
| --- | --- | --- |
| Chữ viết tắt | English | Tiếng việt |
| ISL | Introduction to Statistical Learning | Mô hình dự đoán và phân tích thống kê. |
| HOML | Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow | Machine learning và deep learning sử dụng Scikit-Learn, Keras, và TensorFlow. |
| POA | Production and Operations Analysis | Quản lý sản xuất và phân tích hoạt động sản xuất. |
| SCM | Supply Chain Management: Strategy, Planning, and Operation | Quản lý chuỗi cung ứng và kế hoạch hóa hoạt động. |
| MM | Marketing Management | Quản lý tiếp thị và chiến lược kinh doanh. |
| DDD | Data-Driven: Creating a Data Culture | Tạo ra văn hóa sử dụng dữ liệu trong quyết định kinh doanh. |

|  |  |
| --- | --- |
| Ảnh có chứa văn bản, biểu tượng, Nhãn hiệu, Phông chữ  Mô tả được tạo tự động | **BỘ GIÁO DỤC VÀ ĐÀO TẠO**  **TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH** |

**KHOA CÔNG NGHỆ THÔNG TIN**

**SỔ NHẬT KÝ THỰC TẬP**

Họ tên HSSV: Võ Thái Hiển

Lớp: 20DTH1D

Ngành: Công nghệ thông tin

Cơ quan thực tập

Thời gian thực tập: **Từ …/…/20… đến …/…/20…**

**Tháng …/20…**

**\* Lưu ý:** Nội dung thực tập được liệt kê và đánh giá theo từng tuần trong đợt thực tập.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tuần** | **Thời gian** | **Nội dung thực tập** | **Nhận xét của đơn vị thực tập** | **Chữ ký người HD của đơn vị thực tập** |
| 1 | Từ 09/10/2023  đến 15/10/2023 | Tìm hiểu , nghiên cứu về ngôn ngữ dart ,flutter , cách call API |  |  |
| 2 | Từ 16/10/2023  đến 22/10/2023 | Sử dụng dart, flutter xây dựng ứng dụng Cân bluetooth: streaming |  |  |
| 3 | Từ 23/10/2023  đến 29/10/2023 | Phát triển ứng dụng theo định hướng của doanh nghiệp |  |  |
| 4 | Từ 30/10/2023  đến 5/11/2023 | Phân tích, tìm kiếm dữ liệu có ích cho doanh nghiệp |  |  |
| 5 | Từ 6/11/2023  đến 12/11/2023 | Tìm kiếm quy luật và dữ liệu có ích nhằm tối ưu chi phí sản xuất |  |  |
| 6 | Từ 13/11/2023  đến 19/11/2023 | Ứng dụng các quy luật,dữ liệu có ích đã tìm kiếm vào thực tế |  |  |
| 7 | Từ 20/11/2023  đến 26/11/2023 | Viết báo cáo tổng hợp những gì làm được cho doanh nghiệp |  |  |
| 8 | Từ 27/11/2023  đến 3/12/2023 |  |  |  |

TP.HCM, ngày …...tháng ….. năm 20….

Trưởng đơn vị quản lý thực tập

CHƯƠNG 1./ GIỚI THIỆU CƠ QUAN THỰC TẬP VÀ ĐỀ TÀI

1.1./ Giới thiệu cơ quan thực tập

Ảnh có chứa văn bản, ảnh chụp màn hình, thiết kế

Mô tả được tạo tự động

Hình 1. 1: Hình ảnh của công ty giới thiệu

**Về Công Ty**: Công Ty Bảo Tín Software & Automation 4.0+ được thành lập vào ngày 13/12/2008 và đã trải qua hơn 14 năm phát triển trong ngành "Dệt May – Thực Phẩm – Chế Biến – Năng Lượng". Công Ty tự hào là một trong những đơn vị hàng đầu cung cấp sản phẩm phần mềm và giải pháp tự động hoá cho các ngành công nghiệp đa dạng, từ dệt may, thực phẩm đến chế biến và năng lượng.

**Sứ Mệnh và Giá Trị Cốt Lõi**:Sứ mệnh của Công Ty là cung cấp giải pháp tự động hoá và phần mềm chuyên nghiệp, giúp doanh nghiệp nâng cao hiệu suất công việc, giảm tải nhân sự, tăng chất lượng sản phẩm, giảm chi phí và tăng tính cạnh tranh. Công Ty cam kết đem lại giá trị thực sự cho khách hàng thông qua sự đổi mới và chất lượng dịch vụ.

**Phát Triển và Thành Tích**: Bắt đầu từ năm 2008, Công Ty đã trải qua một hành trình đầy thách thức và đổi mới. Đến năm 2022, chúng tôi đã chuyển đổi thành "Công ty Cổ phần Phần mềm và Công nghệ Tự động 4.0 Bảo Tín", đồng thời trở thành đối tác đáng tin cậy cho hàng trăm doanh nghiệp trong và ngoài nước.

**Sản Phẩm và Dịch Vụ**: Công Ty cung cấp một loạt các sản phẩm phần mềm và giải pháp tự động hoá, bao gồm Quản lý Kho Vật Tư - Phụ Kiện, Quản lý Nhà Máy Sợi - Dệt - Nhuộm, Hệ Thống Cân Chất Hoá Chất Tự Động, và Hệ Thống Máy Kiểm Mộc - Phẩm - Đóng Gói Tự Động. Những sản phẩm này không chỉ giúp tối ưu hóa quá trình sản xuất mà còn nâng cao chất lượng và hiệu suất toàn diện của doanh nghiệp.

**Sản Phẩm**

1. **Quản lý Kho Vật Tư - Phụ Kiện (Catalog 1A-03-01):**

* Mục Tiêu: Cung cấp giải pháp kiểm soát tình trạng tài sản, hỗ trợ quản lý bảo trì và nâng cao hiệu suất tổng thể thiết bị trong thời đại 4.0.
* Chức Năng:
  + Kiểm soát chặt chẽ tình trạng tài sản.
  + Hỗ trợ toàn diện công tác quản lý bảo trì.
  + Tối ưu hóa hiệu suất tổng thể thiết bị.

1. **Quản lý Nhà Máy Sợi (Catalog 2-done-01):**

* Mục Tiêu Chính:
  + Giải pháp hiện đại cho vấn đề cũ.
  + Lập kế hoạch chính xác dựa trên số liệu thực tế về thời gian sản xuất.
  + Tích hợp MRP, Chất lượng, Bảo trì và PLM hoàn toàn.
  + Giao tiếp theo thời gian thực với bảng việc và cảnh báo chất lượng.

1. **Quản lý Nhà Máy Sợi - Dệt - Nhuộm (Catalog 5-02):**

* Mục Tiêu:
  + Kiểm soát hiệu quả quá trình sản xuất.
  + Tích hợp PMS-BT2 giúp nâng cao năng suất.
  + Hỗ trợ các bộ phận trong nhà máy dễ dàng kiểm soát công việc với công nghệ số.

1. **Hệ Thống Cân Chất Hoá Chất Lỏng Tự Động (Catalog 3-01(1)):**

* Mục Tiêu:
  + Tối ưu hóa quá trình sản xuất và quản lý chất liệu.
  + Tiết kiệm thời gian, công sức và chi phí.
  + Ngăn chặn lỗi không nhất thiết phải xảy ra trong quá trình sản xuất.

1. **Hệ Thống Máy Kiểm Mộc - Phẩm - Quấn Đai - Đóng Gói Tự Động (Catalog 6-02):**

* Mục Tiêu:
  + Kiểm soát hàng hóa về chất lượng và số lượng.
  + Quản lý lỗi và đánh giá chất lượng sản phẩm.
  + Hỗ trợ nhiều tính năng tự động như đo số mét, cắt tự động và phân loại hàng hóa.

1. **Hệ Thống Băng Tải Thông Minh - Định Lượng Tự Động (Catalog 7-01):**

* Mục Tiêu:
  + Tăng hiệu suất và hiệu quả công việc.
  + Tối ưu hóa lưu lượng công việc và giảm thời gian làm việc.
  + Cải thiện độ tin cậy và đảm bảo thời gian thực hiện đúng thời hạn.

1. **Phần Mềm Quản Lý Cửa Hàng Thương Mại Ngành Vải (Catalog 8-01):**

* Mục Tiêu:
  + Quản lý nhập xuất hàng từ nhà cung cấp sợi, nhà dệt, nhà gia công nhuộm.
  + Kiểm tra chính xác sản phẩm được xuất bán qua cân điện tử và máy quét mã vạch.
  + Hỗ trợ trên nền tảng mobile IOS và Android.

**Dịch vụ**

**Triển Khai**: Đội ngũ hỗ trợ triển khai tận nơi với kỹ sư phần mềm, điện tử và tự động hoá.

**Hỗ Trợ Khách Hàng:** Đội ngũ hỗ trợ online với 10-15 nhân sự luôn sẵn sàng hỗ trợ khách hàng khi cần.

**Đối Tác Chiến Lược và Quan Hệ Hợp Tác**: Công Ty tự hào về việc có nhiều đối tác chiến lược uy tín trong và ngoài nước, đặc biệt là trong lĩnh vực phần mềm và tự động hoá. Mối quan hệ hợp tác chặt chẽ này giúp Công Ty duy trì sự đổi mới và linh hoạt để đáp ứng nhanh chóng với những thách thức ngày càng đa dạng của thị trường.

**Công Nghệ và R&D**: Công Ty liên tục đầu tư vào nghiên cứu và phát triển để duy trì sự tiên tiến của sản phẩm và dịch vụ của mình. Công nghệ AI, BigData, và Cloud được tích hợp trong các sản phẩm để mang lại trải nghiệm người dùng tốt nhất và đáp ứng nhanh chóng với yêu cầu ngày càng cao của khách hàng.

**Mục Tiêu Phát Triển**: Mục tiêu của Công Ty là không ngừng phát triển và trở thành đối tác tin cậy hàng đầu trong lĩnh vực phần mềm và tự động hoá. Công Ty đặt tầm nhìn vào việc hỗ trợ doanh nghiệp nâng cao khả năng cạnh tranh và thích ứng với môi trường kinh doanh ngày càng biến động.

**Hỗ Trợ và Dịch Vụ**: Công Ty không chỉ cung cấp các sản phẩm chất lượng mà còn đặc biệt chú trọng vào dịch vụ hỗ trợ. Đội ngũ hỗ trợ của Công Ty không ngừng luôn sẵn sàng để giúp đỡ khách hàng từ việc triển khai tới vấn đề kỹ thuật.

**Hình Ảnh Lãnh Đạo**: Ban quản lý của Công Ty dẫn đầu bởi ông Nguyễn Văn Hoà - CEO/Founder, người có tầm nhìn sâu sắc về sự kết hợp giữa công nghệ thông tin và tự động hoá, giúp doanh nghiệp thích ứng và phát triển trong thời đại công nghiệp 4.0.

**Tầm Nhìn Tương Lai**: Công Ty không chỉ nhìn về quá khứ và hiện tại, mà còn đặt tầm nhìn vào tương lai. Với sự đổi mới liên tục và cam kết với chất lượng, Công Ty hy vọng trở thành người đồng hành tin cậy của doanh nghiệp trong hành trình chuyển đổi số và

1.2./ Giới thiệu đề tài

1.2.1./ Lý do chọn đề tài

**Phản Ánh Thực Tế Công Việc:** Đề tài liên quan trực tiếp đến lĩnh vực sản xuất nhuộm, là một phần quan trọng trong hoạt động của công ty. Việc phân tích dữ liệu sản xuất nhuộm mang lại cái nhìn toàn diện về quá trình sản xuất và làm rõ các vấn đề cụ thể trong công việc hàng ngày.

**Ứng Dụng Công Nghệ 4.0:** Đề tài phản ánh sự hợp nhất giữa công nghệ và sản xuất, đặc biệt là trong bối cảnh Công Nghệ 4.0. Phân tích dữ liệu sản xuất nhuộm là cơ hội để áp dụng những công cụ và kỹ thuật hiện đại như trí tuệ nhân tạo, big data để tối ưu hóa quy trình sản xuất.

**Tối Ưu Hóa Hiệu Suất Sản Xuất**:Phân tích dữ liệu giúp xác định những điểm mạnh và yếu trong quy trình sản xuất nhuộm, từ đó đưa ra các biện pháp cụ thể nhằm tối ưu hóa hiệu suất, giảm thiểu lãng phí và tăng cường chất lượng sản phẩm.

**Giải Quyết Vấn Đề Thực Tiễn:** Sản xuất nhuộm là một ngành công nghiệp đầy thách thức với nhiều yếu tố ảnh hưởng đến chất lượng sản phẩm. Phân tích dữ liệu giúp chúng tôi đối mặt và giải quyết các vấn đề thực tiễn một cách khoa học, đồng thời nâng cao khả năng dự đoán và phòng tránh các sự cố.

**Phát Triển Kỹ Năng Phân Tích Dữ Liệu:** Thực hiện đề tài này sẽ giúp tôi phát triển kỹ năng phân tích dữ liệu, từ việc thu thập đến xử lý và đưa ra kết luận. Kỹ năng này không chỉ quan trọng trong ngữ cảnh của công việc hiện tại mà còn là một nền tảng quan trọng cho sự phát triển sự nghiệp trong tương lai.

**Đóng Góp vào Chiến Lược Phát Triển Công Ty:** Kết quả của đề tài sẽ cung cấp thông tin cần thiết để đóng góp vào quá trình đưa ra chiến lược phát triển của công ty, từ việc cải thiện quy trình sản xuất đến tối ưu hóa nguồn lực và gia tăng giá trị cho khách hàng.

**Thực Hiện Dự Án Thực Tế:** Đề tài không chỉ là một bài toán nghiên cứu lý thuyết mà còn là một dự án thực tế, mang lại cơ hội để áp dụng kiến thức học được tại trường vào môi trường làm việc thực tế và định hình sự hiểu biết vững về công việc.

1.2.2./ Mục tiêu của đề tài

**Tối Ưu Hóa Quy Trình Sản Xuất:** Đưa ra biện pháp cụ thể để cải thiện hiệu suất sản xuất nhuộm.Xác định các bước cần được tối ưu hóa để giảm thiểu lãng phí và tăng cường năng suất.

**Dự Đoán và Phòng Tránh Sự Cố:** Phát triển mô hình dự đoán sự cố trong quy trình sản xuất nhuộm.Xây dựng hệ thống cảnh báo để giảm thiểu rủi ro sự cố và giảm tổn thất.

**Nâng Cao Chất Lượng Sản Phẩm:** Phân tích dữ liệu để đảm bảo chất lượng sản phẩm đạt chuẩn.Xác định các yếu tố ảnh hưởng đến chất lượng và đề xuất cải tiến.

**Tăng Cường Sự Linh Hoạt và Tương Tác:** Phát triển giải pháp để tăng cường sự linh hoạt trong sản xuất nhuộm.Tạo ra một hệ thống tương tác linh hoạt với các biến thay đổi trong quy trình.

**Thiết Lập Hệ Thống Báo Cáo Hiệu Quả:** Xây dựng hệ thống báo cáo giúp đánh giá hiệu quả các biện pháp đã thực hiện.Đưa ra các đề xuất và cải thiện dựa trên các kết quả thu được.

1.2.3./ Đối tượng và phạm vi nguyên cứu của đề tài

**Đối tượng nghiên cứu**

Hệ Thống Sản Xuất Nhuộm: Tất cả các quy trình và bước sản xuất nhuộm trong nhà máy.

Dữ Liệu Sản Xuất: Các dữ liệu liên quan đến quá trình sản xuất nhuộm, bao gồm dữ liệu về nguyên liệu, quy trình sản xuất, thời gian, và chất lượng sản phẩm.

Nhân Viên Sản Xuất: Các nhân viên tham gia trực tiếp vào quá trình sản xuất nhuộm.

Hệ Thống Theo Dõi và Đo Lường: Các hệ thống theo dõi và đo lường liên quan đến quy trình sản xuất, bao gồm cả cảm biến và các thiết bị đo lường.

**Phạm vi nghiên cứu:**

Quy Trình Sản Xuất Nhuộm: Phân tích chi tiết từng bước trong quy trình sản xuất nhuộm, từ việc chuẩn bị nguyên liệu đến thành phẩm.

Dữ Liệu Nguồn Gốc: Thu thập và phân tích dữ liệu về nguyên liệu sử dụng trong quá trình sản xuất nhuộm, bao gồm loại và chất lượng.

Chất Lượng Sản Phẩm: Đánh giá chất lượng của sản phẩm nhuộm ở mỗi giai đoạn trong quy trình sản xuất.

Hiệu Suất Sản Xuất: Đánh giá hiệu suất sản xuất thông qua các chỉ số như năng suất, tỷ lệ lãng phí, và thời gian sản xuất.

Dự Đoán và Phòng Tránh Sự Cố: Xây dựng mô hình để dự đoán sự cố có thể xảy ra trong quá trình sản xuất và phát triển các biện pháp phòng ngừa.

Hệ Thống Báo Cáo: Phát triển hệ thống báo cáo giúp theo dõi và đánh giá hiệu quả các biện pháp cải tiến đã được thực hiện.

Nhân Viên Sản Xuất: Đối tượng nghiên cứu sẽ liên quan đến nhân viên sản xuất, đặc biệt là những người làm việc trực tiếp trong quy trình nhuộm.

Hệ Thống Theo Dõi và Đo Lường: Sự tích hợp và phân tích dữ liệu từ các hệ thống theo dõi và đo lường để đảm bảo độ chính xác và toàn vẹn của dữ liệu.

CHƯƠNG 2./ CƠ SỞ LÝ LUẬN VỀ VẤN ĐỀ NGHIÊN CỨU VÀ MÔ HÌNH LÝ THUYẾT

2.1./ Cơ sở lý luận về vấn đề nghiên cứu

**Khái niệm lý thuyết:**

Khách hàng không chỉ là người mua hàng, mà là đối tác chiến lược quan trọng của mọi doanh nghiệp. Mô hình quan hệ với khách hàng không chỉ giới hạn trong quá trình bán hàng, mà còn bao gồm sự tương tác và hỗ trợ sau bán hàng. Để hiểu rõ hơn về họ, phải đặt mình vào vị thế của khách hàng, nắm bắt mong muốn và lo ngại của họ.

Các khía cạnh như đánh giá sản phẩm, trải nghiệm dịch vụ, và tương tác thường xuyên đóng vai trò quan trọng trong quá trình quyết định của khách hàng. Phân tích dữ liệu về hành vi mua hàng, thói quen truy cập trang web, và phản hồi sẽ cung cấp cái nhìn sâu sắc về đối tượng khách hàng.

Các mô hình hành vi của khách hàng như mô hình RFM (Recency, Frequency, Monetary), mô hình AIDA (Awareness, Interest, Desire, Action), hay mô hình Customer Journey đều là những công cụ hữu ích trong việc hiểu rõ hành vi và mong đợi của khách hàng.

Hiểu rõ về khách hàng không chỉ giúp doanh nghiệp tối ưu hóa dịch vụ hiện tại mà còn tạo ra cơ hội để phát triển các sản phẩm và dịch vụ mới phù hợp với nhu cầu tiềm ẩn của họ. Tầm quan trọng của việc xây dựng và duy trì một mối quan hệ chặt chẽ với khách hàng không thể phủ nhận trong bối cảnh cạnh tranh ngày nay, và phân tích dữ liệu chính là chìa khóa để mở cánh cửa này.

**Nghiên cứu trước đó:**

Các nghiên cứu trước đó về phân tích dữ liệu khách hàng đã chứng minh giá trị của việc áp dụng các kỹ thuật này trong thực tế doanh nghiệp. Nghiên cứu của Nguyen et al. (2019) về việc sử dụng phân tích RFM để hiểu rõ hành vi mua sắm và loyalt của khách hàng đã ghi chú đến tăng trưởng doanh số bán hàng và giảm tỷ lệ churn.

Trong một nghiên cứu khác của Smith et al. (2020), việc ứng dụng mô hình Customer Journey Mapping đã dẫn đến sự cải thiện đáng kể trong trải nghiệm của khách hàng và tăng cường sự hiểu biết về hành vi mua hàng.

Nhìn chung, các nghiên cứu trước đó đã thể hiện rằng phân tích dữ liệu khách hàng không chỉ giúp nâng cao sự hiểu biết về đối tượng mục tiêu mà còn mang lại lợi ích kinh tế và chiến lược quan trọng cho tổ chức.

**Nhu cầu nghiên cứu:**

Tại Bảo Tín, nhu cầu tối ưu hóa chiến lược phục vụ khách hàng trở nên ngày càng quan trọng để duy trì và mở rộng thị trường. Nghiên cứu sẽ tập trung vào các khía cạnh sau:

* Tìm Hiểu Sâu Sắc về Khách Hàng: Phân tích dữ liệu để định rõ các nhóm khách hàng, hiểu rõ nhu cầu và mong đợi của từng nhóm.
* Đo Lường Sự Hài Lòng: Xác định các chỉ số để đo lường sự hài lòng của khách hàng và đề xuất biện pháp cải thiện nếu cần thiết.
* Dự Đoán Hành Vi Mua Hàng: Sử dụng mô hình dự đoán để hiểu và dự đoán hành vi mua hàng trong tương lai.

**Lập luận logic:**

Tối ưu hóa chiến lược phục vụ khách hàng thông qua phân tích dữ liệu không chỉ là một lợi ích cho khách hàng mà còn đem lại nhiều lợi ích cho doanh nghiệp:

* Cải Thiện Trải Nghiệm Khách Hàng: Thông qua việc hiểu rõ hơn về hành vi và mong muốn, Bảo Tín có thể cá nhân hóa và cải thiện trải nghiệm của khách hàng.
* Tăng Cường Sự Chân Thực: Các biện pháp cải thiện dựa trên dữ liệu giúp tăng cường sự chân thực và mối quan hệ với khách hàng.
* Tối Ưu Hóa Chiến Lược Tiếp Thị: Phân tích dữ liệu sẽ hỗ trợ trong việc lựa chọn kênh tiếp thị hiệu quả và tối ưu hóa chiến lược quảng cáo.

**Phương pháp nghiên cứu:**

Thu Thập Dữ Liệu: Tập trung vào việc thu thập dữ liệu từ nhiều nguồn như hệ thống CRM, dữ liệu giao dịch, khảo sát khách hàng và phản hồi trực tuyến.

Phân Tích Dữ Liệu: Sử dụng các kỹ thuật phân tích dữ liệu như phân tích đa biến, phân tích cụm, và mô hình học máy để đánh giá và rút ra thông tin quan trọng.

Đánh Giá Kết Quả: Đo lường hiệu suất của chiến lược mới dựa trên các chỉ số kinh doanh quan trọng như doanh số bán hàng, sự hài lòng của khách hàng, và tỷ lệ giữ chân khách hàng.

2.2./ Cơ sở lý thuyết

2.2.1/ Thu thập dữ liệu

Thu thập dữ liệu là quá trình quan trọng và phức tạp trong nghiên cứu khoa học và phân tích dữ liệu. Để đảm bảo tính chính xác, đáng tin cậy và hữu ích của dữ liệu, cần phải tuân thủ các nguyên tắc và phương pháp đúng đắn trong quá trình thu thập. Dưới đây là một lý thuyết về thu thập dữ liệu:

**Xác Định Mục Tiêu Thu Thập Dữ Liệu:** Trước khi bắt đầu quá trình thu thập dữ liệu, việc xác định mục tiêu rõ ràng là quan trọng để định hình chiến lược nghiên cứu. Mục tiêu này giúp xác định phạm vi và tính chính xác của dữ liệu cần thu thập.

**Chọn Phương Pháp Thu Thập Dữ Liệu:** Có nhiều phương pháp khác nhau để thu thập dữ liệu, và lựa chọn phương pháp phụ thuộc vào bản chất của mục tiêu nghiên cứu. Khảo sát, phỏng vấn, quan sát, thử nghiệm, và phân tích nội dung là những phương pháp phổ biến được áp dụng.

**Xây Dựng Kế Hoạch Thu Thập Dữ Liệu:** Kế hoạch thu thập dữ liệu cần được thiết kế cụ thể, bao gồm việc xác định nguồn dữ liệu, đối tượng mẫu, phương thức thu thập, và lên lịch trình. Kế hoạch này giúp đảm bảo rằng quá trình thu thập diễn ra mạch lạc và hiệu quả.

**Lựa Chọn Đối Tượng Mẫu:** Việc chọn đối tượng mẫu quan trọng để đảm bảo tính đại diện của nghiên cứu. Phải xác định rõ đặc điểm của đối tượng mẫu để kết quả có thể áp dụng rộng rãi và đáng tin cậy.

**Đảm Bảo Chất Lượng Dữ Liệu:** Chất lượng dữ liệu ảnh hưởng trực tiếp đến độ chính xác của kết quả nghiên cứu. Cần thiết lập các biện pháp kiểm soát chất lượng, đào tạo nhân viên thu thập dữ liệu, và sử dụng công cụ thu thập chính xác.

**Phát Hiện và Xử Lý Sai Số:** Trong quá trình thu thập, sai số có thể xuất hiện. Việc phát hiện và xử lý sai số là quan trọng để đảm bảo tính chính xác và đáng tin cậy của dữ liệu thu thập được.

**Bảo Vệ Quyền Riêng Tư và Đạo Đức:** Việc bảo vệ quyền riêng tư của người tham gia nghiên cứu là không thể phủ nhận. Cần phải tuân thủ các nguyên tắc đạo đức và quy định pháp luật liên quan đến bảo vệ thông tin cá nhân.

**Hiểu Rõ Ngữ Cảnh:** Để giải thích và diễn giải dữ liệu, cần phải hiểu rõ ngữ cảnh trong đó dữ liệu được thu thập. Điều này bao gồm việc nắm vững môi trường, đối tượng nghiên cứu, và yếu tố có thể ảnh hưởng đến dữ liệu.

**Đánh Giá Độ Tin Cậy và Thực Thi:** Trước khi sử dụng dữ liệu trong phân tích, cần phải đánh giá độ tin cậy của nó. Kiểm tra độ đầy đủ, chính xác, và khả năng tổng quát của dữ liệu giúp đảm bảo rằng kết quả là đáng tin cậy và có giá trị.

2.2.2./ Tiền xử lý dữ liệu

Tiền xử lý dữ liệu (Data Preprocessing) là quá trình chuẩn bị và làm sạch dữ liệu để nó có thể được hiệu quả sử dụng trong các tác vụ phân tích dữ liệu hoặc huấn luyện mô hình máy học. Dưới đây là một số khái niệm lý thuyết quan trọng trong tiền xử lý dữ liệu:

Loại bỏ Dữ liệu Ngoại lai (Outliers): Outliers (Ngoại lai): Là các điểm dữ liệu có giá trị rất lớn hoặc rất nhỏ so với phần còn lại của dữ liệu.

Phương pháp xử lý: Sử dụng biểu đồ như boxplot để phát hiện outliers. Có thể quyết định loại bỏ hoặc thay thế outliers dựa trên phân phối dữ liệu và kiến thức chuyên môn.

Xử lý Giá trị Thiếu (Missing Values): Giá trị Thiếu (Missing Values): Là các ô dữ liệu mà không có giá trị hoặc được đánh dấu là NaN (Not a Number).

Phương pháp xử lý: Loại bỏ dòng hoặc cột chứa giá trị thiếu. Thay thế giá trị thiếu bằng giá trị trung bình, trung vị hoặc dự đoán từ mô hình.

Chuẩn hóa Dữ liệu (Data Normalization): Chuẩn hóa: Là quá trình chuyển đổi giá trị của biến sao cho chúng nằm trong một khoảng cụ thể.

Phương pháp chuẩn hóa: Standardization: Chia giá trị của biến cho độ lệch chuẩn (standard deviation). Min-Max Scaling: Chuyển đổi giá trị của biến sao cho chúng nằm trong khoảng [0, 1]. Mã hóa Biến Phân loại (Categorical Variable Encoding): Mã hóa one-hot (One-Hot Encoding): Tạo ra các cột mới để biểu diễn mỗi giá trị của biến phân loại. Mã hóa số (Label Encoding): Gán một số duy nhất cho mỗi giá trị của biến phân loại.

Kiểm tra và Chọn Tính năng (Feature Selection): Tính năng (Features): Là các biến hoặc thuộc tính trong dữ liệu.

Phương pháp chọn tính năng: Recursive Feature Elimination (RFE): Loại bỏ các tính năng ít quan trọng lẫn lượt.

Importance from Trees: Xác định độ quan trọng của từng tính năng từ các mô hình cây quyết định.

Giảm Kích thước Dữ liệu (Dimensionality Reduction): Kích thước dữ liệu: Là số lượng biến trong dữ liệu.

Phương pháp giảm kích thước: Principal Component Analysis (PCA): Chuyển đổi dữ liệu thành các thành phần chính (principal components) để giảm kích thước.

Làm Mềm Dữ liệu (Data Imputation): Làm mềm (Imputation): Là quá trình dự đoán và điền vào giá trị thiếu.

Phương pháp làm mềm: Sử dụng mô hình học máy để dự đoán giá trị thiếu dựa trên các biến khác.

2.2.3./ Phân tích dữ liệu

Phân tích dữ liệu là quá trình chuyển đổi thông tin không có trật tự thành kiến thức có ý nghĩa, giúp hiểu rõ hơn về các xu hướng, mối quan hệ, và sự biến đổi trong dữ liệu. Dưới đây là một lý thuyết về phân tích dữ liệu:

**Xác Định Mục Tiêu Phân Tích**: Mục tiêu phân tích phải được xác định rõ ràng từ ban đầu. Có thể là tìm hiểu xu hướng, dự đoán tương lai, hoặc đơn giản là mô tả dữ liệu. Mục tiêu này sẽ định hình phương pháp và kỹ thuật phân tích sử dụng.

**Chọn Phương Pháp Phân Tích Thích Hợp:** Có nhiều phương pháp phân tích dữ liệu như phân tích số liệu mô tả, phân tích hồi quy, phân tích nhóm, phân tích chuỗi thời gian, và phân tích phân cụm. Lựa chọn phương pháp phụ thuộc vào loại dữ liệu và mục tiêu cụ thể.

**Kiểm Tra Và Chuẩn Bị Dữ Liệu:** Trước khi bắt đầu phân tích, cần kiểm tra và chuẩn bị dữ liệu. Điều này bao gồm loại bỏ giá trị thiếu, kiểm tra tính đồng nhất của dữ liệu, và biến đổi dữ liệu nếu cần thiết.

**Thực Hiện Phân Tích:** Thực hiện các bước phân tích dữ liệu theo phương pháp đã chọn. Sử dụng các công cụ và phần mềm phân tích thống kê như R, Python, hoặc các công cụ chuyên nghiệp như SPSS và SAS.

**Hiểu Rõ Ngữ Cảnh:** Hiểu rõ ngữ cảnh của dữ liệu là quan trọng. Điều này bao gồm việc nắm vững kiến thức chuyên sâu về lĩnh vực cụ thể của dữ liệu để có thể diễn giải kết quả phân tích một cách đúng đắn.

**Diễn Giải Kết Quả:** Kết quả phân tích cần được diễn giải một cách có ý nghĩa với mục tiêu nghiên cứu. Điều này bao gồm việc xác định các xu hướng, mối quan hệ, và ý nghĩa thống kê của các kết quả.

**Tổng Hợp Kết Quả và Đưa Ra Kết Luận:** Tổng hợp các kết quả phân tích để đưa ra kết luận và giải đáp cho các câu hỏi nghiên cứu. Nêu rõ những điểm mạnh và yếu của phân tích để tạo nền tảng cho các bước nghiên cứu tiếp theo.

**Kiểm Định Độ Tin Cậy Của Kết Quả:** Để đảm bảo tính chính xác của kết quả, cần kiểm tra độ tin cậy thông qua việc sử dụng các phương pháp thống kê như kiểm định t, ANOVA, hoặc các phương pháp khác tùy thuộc vào loại phân tích được thực hiện.

2.2.4./ Tổng quan về Regression trong Machine Learning

2.2.4.1./ Linear Regression (Hồi Quy Tuyến Tính)

Linear Regression (Hồi Quy Tuyến Tính) là một phương pháp trong Machine Learning và thống kê được sử dụng để mô hình hóa mối quan hệ tuyến tính giữa biến độc lập và biến phụ thuộc. Đây là một trong những phương pháp đơn giản nhất và phổ biến nhất trong học máy.

Các Đặc Điểm Cơ Bản của Linear Regression:

* Biến Phụ Thuộc (Dependent Variable): Là biến mà chúng ta muốn dự đoán. Trong Linear Regression, biến này thường là liên tục, ví dụ như giá cả, doanh số bán hàng, v.v.
* Biến Độc Lập (Independent Variable): Là các biến được sử dụng để dự đoán biến phụ thuộc. Chúng có thể là một hoặc nhiều biến, và thường là dữ liệu số.
* Mô Hình Hồi Quy Tuyến Tính: Mô hình dựa trên giả định rằng mối quan hệ giữa các biến độc lập và biến phụ thuộc là tuyến tính. Trong trường hợp đơn giản nhất với một biến độc lập, mô hình có dạng:

*y = mx + b*

Trong đó, *y* là biến phụ thuộc, *x* là biến độc lập, *m* là độ dốc của đường thẳng (slope), và *b* là điểm cắt trục tung (intercept).

Quá Trình Huấn Luyện Mô Hình:

* Tìm Đường Thẳng Tốt Nhất: Mục tiêu của huấn luyện là tìm ra đường thẳng (hoặc siêu mặt phẳng trong trường hợp đa chiều) sao cho sai số dự đoán của mô hình là nhỏ nhất.
* Phương Pháp Tối Ưu Hóa: Thông thường, Linear Regression sử dụng phương pháp tối ưu hóa như Ordinary Least Squares (OLS) để tìm ra các giá trị \(m\) và \(b\) tối ưu nhất.

Đánh Giá Mô Hình:

* Mean Squared Error (MSE) và RMSE: Dùng để đánh giá hiệu suất của mô hình. MSE đo lường trung bình của bình phương sai số giữa dự đoán và giá trị thực tế, trong khi RMSE là căn bậc hai của MSE.
* R-squared (R²): Đo lường mức độ biến thiên của biến phụ thuộc được giải thích bởi mô hình. Giá trị R² càng gần 1 thể hiện mô hình giải thích tốt dữ liệu.

2.2.4.2./ Random Forest Regression (Rừng Ngẫu Nhiên Hồi Quy)

Random Forest Regression là một phương pháp học máy trong đó một loạt các cây quyết định (decision trees) được tạo ra và kết hợp lại để tạo thành một mô hình dự đoán. Khác với việc sử dụng một cây quyết định duy nhất, Random Forest sử dụng nhiều cây và kết hợp kết quả từ các cây này để đưa ra dự đoán chính xác hơn.

Đặc Điểm Cơ Bản của Random Forest Regression:

* Tạo Ra Nhiều Cây Quyết Định (Decision Trees): Random Forest tạo ra một tập hợp các cây quyết định độc lập nhau, mỗi cây được huấn luyện trên một tập dữ liệu con và một số lượng nhỏ các đặc trưng được chọn ngẫu nhiên.
* Phương Pháp Bầu Cử (Voting Method): Khi có dữ liệu mới cần dự đoán, mỗi cây trong Random Forest sẽ đưa ra dự đoán của mình, và kết quả cuối cùng sẽ dựa trên việc "bầu cử" từ tất cả các cây để đưa ra kết quả cuối cùng.
* Ứng Dụng Trong Hồi Quy: Trong trường hợp của Regression, Random Forest Regression tính trung bình kết quả của tất cả các cây để đưa ra dự đoán.

Điểm Mạnh và Ứng Dụng:

* Khả Năng Xử Lý Dữ Liệu Lớn: Random Forest có khả năng xử lý các tập dữ liệu lớn và có nhiều đặc trưng mà không cần nhiều tiền xử lý dữ liệu.
* Chống Overfitting (Quá Mức Khớp Dữ Liệu): Bởi vì kết quả cuối cùng được tính từ nhiều cây khác nhau, nó thường ít bị overfitting hơn so với một cây quyết định duy nhất.
* Ứng Dụng Rộng Rãi: Random Forest có thể được áp dụng trong nhiều bài toán hồi quy như dự đoán giá nhà, dự đoán doanh thu, hoặc các bài toán có dữ liệu số lượng lớn.

Random Forest Regression là một phương pháp mạnh mẽ và linh hoạt trong việc dự đoán giá trị liên tục dựa trên dữ liệu đa chiều. Sự kết hợp của nhiều cây quyết định giúp cải thiện độ chính xác và ổn định của mô hình.

CHƯƠNG 3./ MÔ HÌNH THỰC NGHIỆM

3.1./ Bộ dữ liệu

3.1.1/ Bộ dữ liệu Chemical Using

Bộ dữ liệu “Chemical Using” là một bảng chứa thông tin về sử dụng hóa chất, với các cột như ChemicalId, Kg, RealKg, ScalerId, StartScale và EndScale. Dưới đây là một số phân tích cơ bản mà bạn có thể thực hiện trên bộ dữ liệu này:

**Tổng quan:**

Số lượng dòng dữ liệu: 14666 dòng.

Các cột: ChemicalId, Kg, RealKg, ScalerId, StartScale, EndScale.

**Thống kê mô tả:** Sử dụng các hàm như mean, median, min, max để hiểu về phân phối của dữ liệu trong từng cột số.

**Sử dụng hóa chất:**

ChemicalId là một định danh cho loại hóa chất. Bạn có thể xem xét tần suất xuất hiện của mỗi ChemicalId để hiểu loại hóa chất nào được sử dụng nhiều nhất.

Phân tích Kg và RealKg để so sánh lượng hóa chất được kế hoạch sử dụng (Kg) và lượng thực tế sử dụng (RealKg).

**ScalerId:** ScalerId là một định danh cho bộ chia tỷ lệ. Bạn có thể xem xét tần suất xuất hiện của mỗi ScalerId để hiểu tỷ lệ nào được sử dụng nhiều nhất.

**Thời gian sử dụng:** StartScale và EndScale là thời điểm bắt đầu và kết thúc sử dụng hóa chất. Bạn có thể xem xét khoảng thời gian sử dụng để hiểu thời gian sử dụng trung bình, thời điểm sử dụng phổ biến, v.v.

**Phân tích đặc biệt:**

Kiểm tra liệu có bất kỳ giá trị lạ nào hoặc giá trị bị thiếu trong dữ liệu không.

Nếu có dữ liệu liên tục về sử dụng hóa chất, bạn có thể xem xét xu hướng theo thời gian.

3.1.2./ Bộ dữ liệu Sale

Bộ dữ liệu “Sale” liên quan đến các sản phẩm và giao dịch, với các cột như ProductId, TotalQuantity, Quantity, CreatedDate và CustomerId. Dưới đây là một số phân tích cơ bản mà bạn có thể thực hiện trên bộ dữ liệu này:

**Tổng quan:**

Số lượng dòng dữ liệu: 10567 dòng.

Các cột: ProductId, TotalQuantity, Quantity, CreatedDate, CustomerId.

**Thống kê mô tả:** Sử dụng các hàm như mean, median, min, max để hiểu về phân phối của dữ liệu trong từng cột số (TotalQuantity, Quantity).

**Số lượng sản phẩm (ProductId):** Xem xét số lượng sản phẩm khác nhau được bán hoặc mua.

**Tổng số lượng và tổng lượng sản phẩm:** Tổng Quantity và TotalQuantity để hiểu tổng lượng và số lượng sản phẩm đã được bán hoặc mua.

**Thời gian giao dịch (CreatedDate):**

Điều chỉnh định dạng thời gian để có thể phân tích dễ dàng hơn (ví dụ: chuyển đổi sang số phút hoặc giây).

Phân tích xu hướng thời gian, xem xét các khoảng thời gian phổ biến cho giao dịch sản phẩm.

**Khách hàng (CustomerId):** Xem xét số lượng giao dịch từng khách hàng để hiểu khách hàng nào đang mua nhiều nhất hoặc thường xuyên mua hàng.

**Phân tích đặc biệt:**

Kiểm tra liệu có bất kỳ giá trị lạ nào hoặc giá trị bị thiếu trong dữ liệu không.

Nếu có dữ liệu liên tục về giao dịch sản phẩm, bạn có thể xem xét xu hướng theo thời gian.

3.2./ Mô hình thực nghiệm

Trong thời gian thực tập tại công ty, nhiệm vụ của tôi là thực hiện phân tích dữ liệu "Chemical Using" và "Sale" với sự tập trung đặc biệt vào khía cạnh công nghệ thông tin. Cả hai bộ dữ liệu này cung cấp thông tin quan trọng về quy trình sản xuất và doanh số bán hàng, và mục tiêu của tôi là tối ưu hóa việc sử dụng công nghệ trong quá trình phân tích.

**Mục tiêu:**

* Hiểu rõ cách tích hợp và tối ưu hóa quy trình phân tích dữ liệu sử dụng hóa chất và doanh số bán hàng thông qua các công nghệ mới nhất.
* Phát triển và triển khai các giải pháp phần mềm để tự động hóa việc xử lý dữ liệu, giúp giảm thiểu thời gian và tăng độ chính xác.
* Nghiên cứu và áp dụng các công nghệ AI và Machine Learning để dự đoán và tối ưu hóa mô hình sử dụng hóa chất và dự báo doanh số bán hàng.

**Phương pháp thực hiện:**

* Sử dụng ngôn ngữ lập trình và công cụ phân tích dữ liệu để xử lý, biến đổi, và trực quan hóa thông tin một cách hiệu quả.
* Tìm kiếm cách tích hợp các hệ thống IT hiện tại để tối ưu hóa việc thu thập và xử lý dữ liệu.
* Xây dựng và triển khai các ứng dụng phần mềm, dashboard, hoặc các công cụ tự động hóa quy trình phân tích.

3.2.1./ Bộ dữ liệu Chemical Using

Tiến hành thực hiện một số thao tác cơ bản trên bộ dữ liệu "Chemical Using.csv" và có thể mang lại một số thông tin hữu ích

Ảnh có chứa văn bản, phần mềm, Phần mềm đa phương tiện, ảnh chụp màn hình

Mô tả được tạo tự động

Hình 3. 1: Ví dụ code thao tác cơ bản trên bộ dữ liệu

**import pandas as pd**: Nhập thư viện pandas để làm việc với dữ liệu dưới dạng DataFrame, giúp thuận tiện trong việc xử lý và phân tích dữ liệu.

**df = pd.read\_csv('/content/drive/MyDrive/Thực Tập/Chemical Using.csv'):** Đọc dữ liệu từ tệp CSV "Chemical Using.csv" và lưu vào DataFrame có tên là df. Điều này giúp chuẩn bị dữ liệu cho việc phân tích.

**df = df.reset\_index():** Đặt lại chỉ số (index) của DataFrame, có thể giúp đảm bảo chỉ số đang được sử dụng là duy nhất và liên tục.

**df['Re\_Kg\_RealKg'] = df['Kg'] - df['RealKg']:** Tạo một cột mới có tên là 'Re\_Kg\_RealKg', thể hiện sự chênh lệch giữa cột 'Kg' và 'RealKg'. Điều này có thể mang lại thông tin về mức độ chênh lệch giữa dự định sử dụng và thực tế sử dụng hóa chất.

**print(df.head()):** In ra năm dòng đầu tiên của DataFrame sau khi thêm cột mới. Điều này giúp kiểm tra xem cột mới đã được thêm thành công hay không.

**print(df['Re\_Kg\_RealKg'].sum()):** In tổng của cột mới 'Re\_Kg\_RealKg'. Điều này có thể mang lại cái nhìn tổng quan về tổng lượng chênh lệch giữa 'Kg' và 'RealKg' trên toàn bộ bộ dữ liệu.

Đoạn code này có thể cung cấp thông tin về sự chênh lệch giữa lượng hóa chất dự định và lượng thực tế được sử dụng. Thông tin này có thể hữu ích trong việc quản lý nguồn lực, đánh giá hiệu suất và đề xuất các biện pháp tối ưu hóa trong việc sử dụng hóa chất, giúp giảm chi phí và tối ưu hóa quy trình sản xuất.

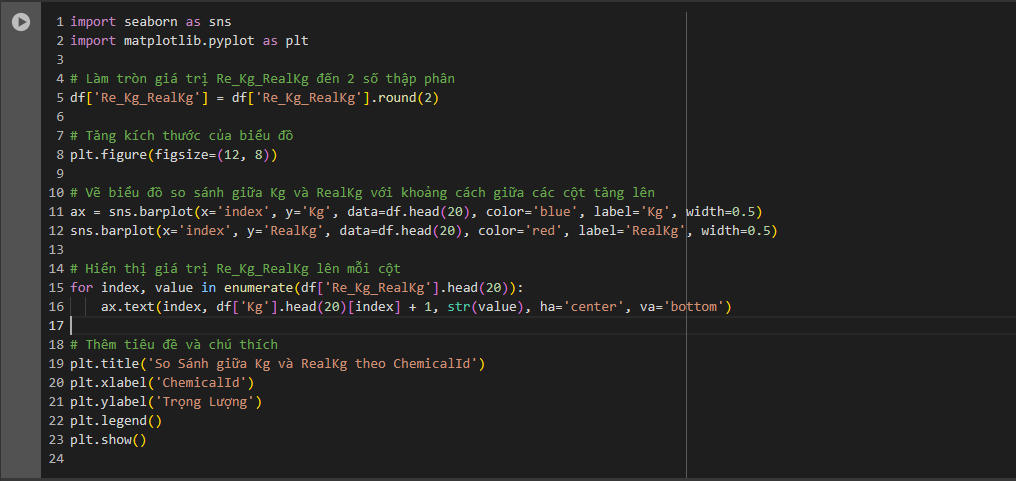
Output:

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, đơn sắc

Mô tả được tạo tự động

Hình 3. 2: Kết quả của thao tác cơ bản

Tiếp theo em viết 1 đoạn code tạo một biểu đồ so sánh giữa lượng 'Kg' và 'RealKg' cho 20 dòng đầu tiên của DataFrame. Ngoài ra, giá trị của cột 'Re\_Kg\_RealKg' được hiển thị lên mỗi cột để thể hiện sự chênh lệch giữa 'Kg' và 'RealKg'



Hình 3. 3: Code tạo biểu đồ so sánh

**df['Re\_Kg\_RealKg'] = df['Re\_Kg\_RealKg'].round(2):** Làm tròn giá trị của cột 'Re\_Kg\_RealKg' đến 2 số thập phân, giúp làm cho biểu đồ trở nên dễ đọc hơn.

plt.figure(figsize=(12, 8)): Thiết lập kích thước của biểu đồ là 12x8 inches, giúp biểu đồ trở nên rõ ràng và dễ nhìn.

**ax = sns.barplot(x='index', y='Kg', data=df.head(20), color='blue', label='Kg', width=0.5):** Sử dụng thư viện seaborn để vẽ biểu đồ cột. Dữ liệu của cột 'Kg' cho 20 dòng đầu tiên được hiển thị bằng cột màu xanh, và 'index' là trục x.

**sns.barplot(x='index', y='RealKg', data=df.head(20), color='red', label='RealKg', width=0.5):** Tương tự như trên, nhưng dữ liệu của cột 'RealKg' được hiển thị bằng cột màu đỏ.

**for index, value in enumerate(df['Re\_Kg\_RealKg'].head(20)): ...:** Dùng vòng lặp để hiển thị giá trị của cột 'Re\_Kg\_RealKg' lên mỗi cột tương ứng của biểu đồ.

**plt.title('So Sánh giữa Kg và RealKg theo ChemicalId'):** Đặt tiêu đề cho biểu đồ.

**plt.xlabel('ChemicalId') và plt.ylabel('Trọng Lượng'):** Đặt nhãn cho trục x và trục y, giúp người đọc hiểu rõ hơn về ý nghĩa của các giá trị trên biểu đồ.

**plt.legend():** Hiển thị chú thích cho biểu đồ để người đọc có thể nhận biết được mỗi cột là 'Kg' hoặc 'RealKg'.

**plt.show():** Hiển thị biểu đồ.

Đoạn code này giúp công ty có cái nhìn tổng quan về sự chênh lệch giữa lượng 'Kg' (dự định) và 'RealKg' (thực tế) cho từng ChemicalId. Biểu đồ có thể giúp trong việc đánh giá hiệu suất sản xuất và quản lý lượng hóa chất sử dụng, đồng thời làm nổi bật các trường hợp có sự chênh lệch lớn giữa dự định và thực tế.

Output:

Ảnh có chứa văn bản, ảnh chụp màn hình, Sơ đồ, biểu đồ

Mô tả được tạo tự động

Hình 3. 4: Biểu đồ so sánh

Tiếp theo em sử dụng thư viện RandomForestRegressor để tạo một mô hình dự đoán

Bước đầu tiên thực hiện một số thao tác để chuyển đổi các cột 'StartScale' và 'EndScale' sang kiểu dữ liệu datetime, sau đó thêm một cột mới có tên là 'TimeDifference' để lưu trữ khoảng thời gian giữa 'EndScale' và 'StartScale'.

Ảnh có chứa văn bản, phần mềm, Phần mềm đa phương tiện, ảnh chụp màn hình

Mô tả được tạo tự động

Hình 3. 5: Thao tác đầu tiên cho mô hình dự đoán

**df['StartScale'] = pd.to\_datetime(df['StartScale']) và df['EndScale'] = pd.to\_datetime(df['EndScale']):** Sử dụng hàm pd.to\_datetime() của thư viện pandas để chuyển đổi cột 'StartScale' và 'EndScale' từ định dạng chuỗi sang kiểu dữ liệu datetime. Điều này giúp hiểu rõ hơn về thời gian và thực hiện các thao tác thời gian.

**df['TimeDifference'] = (df['EndScale'] - df['StartScale']):** Tạo một cột mới có tên là 'TimeDifference' để lưu trữ sự chênh lệch thời gian giữa 'EndScale' và 'StartScale'. Khi bạn trừ hai cột datetime, kết quả sẽ là một đối tượng timedelta, tức là một khoảng thời gian.

**print(df.head()):** In ra năm dòng đầu tiên của DataFrame sau khi thêm cột mới.

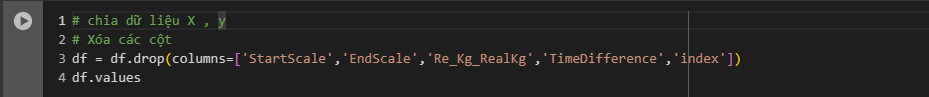
Output:

Ảnh có chứa văn bản, ảnh chụp màn hình

Mô tả được tạo tự động

Hình 3. 6: Kết quả cột datatime mới

Tiếp theo thực hiện một số thao tác để chuẩn bị dữ liệu cho việc dự đoán giá trị cột 'RealKg'.



Hình 3. 7: Xoá các cột không cần thiết

**Df=df.drop(columns=['StartScale','EndScale','Re\_Kg\_RealKg','TimeDifference','index']):** Sử dụng phương thức drop của pandas để loại bỏ các cột không cần thiết khỏi DataFrame. Cụ thể, các cột được liệt kê trong danh sách **['StartScale','EndScale','Re\_Kg\_RealKg','TimeDifference','index']** sẽ bị xóa khỏi DataFrame. Điều này có thể giúp tăng hiệu suất và giảm số lượng dữ liệu không cần thiết khi thực hiện dự đoán.

**df.values:** Truy cập giá trị của DataFrame dưới dạng mảng numpy. Một số mô hình máy học yêu cầu đầu vào là mảng numpy thay vì DataFrame, và bước này giúp chuyển đổi DataFrame thành một mảng numpy.

Output:

Ảnh có chứa văn bản, ảnh chụp màn hình, màu đen, màn hình

Mô tả được tạo tự động

Hình 3. 8: Kết quả sau khi xoá

Tiếp theo sử dụng thư viện scikit-learn để thực hiện việc chia dữ liệu thành tập huấn luyện và tập kiểm tra để chuẩn bị cho quá trình đào tạo và kiểm thử mô hình học máy.

Ảnh có chứa văn bản, Phần mềm đa phương tiện, phần mềm, ảnh chụp màn hình

Mô tả được tạo tự động

Hình 3. 9: Chia dữ liệu cho cột x,y

**from sklearn.model\_selection import train\_test\_split:** Nhập hàm train\_test\_split từ scikit-learn, được sử dụng để chia dữ liệu thành các tập huấn luyện và tập kiểm tra.

**X = df[['ChemicalId', 'Kg', 'ScalerId']]:** Xác định các đặc trưng (features) được sử dụng để đào tạo mô hình. Trong trường hợp này, các đặc trưng là 'ChemicalId', 'Kg', và 'ScalerId'.

**y = df['RealKg']:** Xác định biến mục tiêu (target variable) là 'RealKg', tức là giá trị mà mô hình sẽ cố gắng dự đoán.

**train\_test\_split(X, y, test\_size=0.2, random\_state=42):** Sử dụng hàm train\_test\_split để chia dữ liệu thành tập huấn luyện và tập kiểm tra. Cụ thể:

**X:** Là ma trận chứa các đặc trưng.

**y:** Là vector chứa biến mục tiêu.

**test\_size=0.2:** Xác định tỷ lệ dữ liệu sẽ được chia vào tập kiểm tra. Ở đây, 20% của dữ liệu sẽ được chia vào tập kiểm tra, và 80% còn lại sẽ được sử dụng cho việc huấn luyện mô hình.

**Random\_state=42**: Thiết lập một giá trị seed để đảm bảo rằng kết quả chia dữ liệu là nhất quán khi code được chạy lại.

**X\_train, X\_test, y\_train, y\_test:** Các biến này lưu trữ kết quả của hàm train\_test\_split. X\_train và y\_train là tập huấn luyện, còn X\_test và y\_test là tập kiểm tra. Các tập dữ liệu này sẽ được sử dụng để huấn luyện và kiểm thử mô hình, giúp đánh giá hiệu suất của mô hình trên dữ liệu mới mà nó chưa thấy.

Tiếp theo sử dụng thư viện scikit-learn để chọn và huấn luyện một mô hình học máy để dự đoán giá trị 'RealKg'.

Ảnh có chứa văn bản, Phông chữ, phần mềm, ảnh chụp màn hình

Mô tả được tạo tự động

Hình 3. 10: Code huấn luyện mô hình

**from sklearn.linear\_model import LinearRegression và from sklearn.ensemble import RandomForestRegressor:** Nhập hai loại mô hình từ scikit-learn. LinearRegression là mô hình hồi quy tuyến tính, trong khi RandomForestRegressor là mô hình rừng ngẫu nhiên sử dụng cho hồi quy.

**model = RandomForestRegressor():** Chọn mô hình hồi quy rừng ngẫu nhiên để sử dụng trong việc dự đoán giá trị 'RealKg'. Bạn cũng có thể thay thế bằng LinearRegression() nếu muốn sử dụng mô hình hồi quy tuyến tính.

model.fit(X\_train, y\_train): Huấn luyện mô hình sử dụng tập huấn luyện (X\_train là ma trận các đặc trưng và y\_train là vector biến mục tiêu). Trong trường hợp này, mô hình sẽ học cách dự đoán 'RealKg' dựa trên các đặc trưng 'ChemicalId', 'Kg', và 'ScalerId'.

**y\_pred = model.predict(X\_test):** Dự đoán giá trị 'RealKg' trên tập kiểm tra (X\_test là ma trận các đặc trưng của tập kiểm tra). Kết quả sẽ được lưu vào y\_pred, là vector chứa giá trị dự đoán.

Như vậy là em đã hoàn thành xong mô hình dự đoán bây giờ em sẽ tiến hành kiểm tra thử xem kết quả có đúng như mình muốn không

Ảnh có chứa văn bản, Phần mềm đa phương tiện, phần mềm, ảnh chụp màn hình

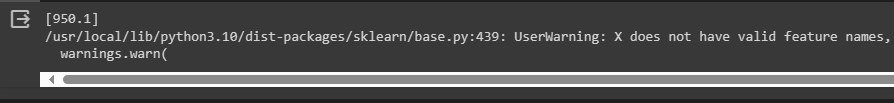
Mô tả được tạo tự động

Hình 3. 11: Kiểm tra kết quả của mô hình

**data1 = 58, data2 = 999, data3 = 1041:** Nhập giá trị cho các đặc trưng 'ChemicalId', 'Kg', và 'ScalerId' của một điểm dữ liệu mới mà bạn muốn mô hình dự đoán.

**print(model.predict([[data1,data2,data3]])):** Sử dụng mô hình đã được huấn luyện để dự đoán giá trị 'RealKg' cho điểm dữ liệu mới này. Kết quả được in ra màn hình.

Output:



Hình 3. 12: Kết quả kiểm tra mô hình

Kết quả dự đoán của mô hình cho điểm dữ liệu mới (ChemicalId: 58, Kg: 999, ScalerId: 1041) là khoảng 950.1. Lưu ý cảnh báo (UserWarning) xuất hiện về việc các đặc trưng không có tên đặc trưng hợp lệ. Cảnh báo này không ảnh hưởng đáng kể đến kết quả dự đoán cụ thể trong trường hợp này.

Mô hình dự đoán này có thể mang lại một số lợi ích cho công ty tùy thuộc vào ngữ cảnh sử dụng

Dự đoán Lượng Hóa Chất Sử Dụng: Mô hình có thể được sử dụng để dự đoán lượng hóa chất sẽ được sử dụng dựa trên thông tin về ChemicalId, Khối lượng (Kg), và ScalerId. Điều này giúp công ty ước tính và quản lý lượng hóa chất cần thiết cho các quy trình sản xuất một cách hiệu quả.

Tối Ưu Hóa Nguyên Liệu và Chi Phí: Dự đoán chính xác về lượng hóa chất cần thiết có thể giúp tối ưu hóa sử dụng nguyên liệu và giảm chi phí sản xuất.

Lên Lịch và Quản Lý Quy Trình: Công ty có thể sử dụng dự đoán để lên lịch sản xuất và quản lý quy trình một cách thông minh, dựa trên dự báo về lượng hóa chất cần thiết.

Giảm Lãng Phí và Rủi Ro: Điều này giúp giảm lãng phí bằng cách đảm bảo sử dụng lượng hóa chất phù hợp và giảm rủi ro về mất mát hoặc thiếu hụt trong quá trình sản xuất.

Tối Ưu Hóa Hiệu Quả Sử Dụng Hóa Chất: Dự đoán chính xác giúp tối ưu hóa hiệu quả sử dụng hóa chất, giảm tác động đến môi trường và tuân thủ các quy định về an toàn và bảo vệ môi trường.

Dự Báo Nhu Cầu Tương Lai: Công ty có thể sử dụng mô hình để dự báo nhu cầu tương lai về hóa chất dựa trên xu hướng lịch sử và các yếu tố thay đổi.

3.2.2./ Bộ dữ liệu Sale

Còn với bộ dữ liệu “Sale” em sẽ tiến hành thống kê và trực quan hoá dữ liệu

Đầu tiên em sẽ sử dụng code để tiến hành thông kê xem có bao nhiêu id sản phẩm được bán ra

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Phần mềm đa phương tiện

Mô tả được tạo tự động

Hình 3. 13: Thông kê id sản phẩm bán ra

Sau đó lọc sản phẩm có số lượng bán cao nhất và thấp nhất từ bộ dữ liệu sale. Tuy nhiên, nếu em có một cột khác là ProductId trong DataFrame, em có thể muốn hiển thị thông tin về sản phẩm đó để có cái nhìn tổng quan hơn.

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, màn hình

Mô tả được tạo tự động

Hình 3. 14: Code thống kê sản phẩm bán chạy nhất và thấp nhất

**max\_total\_quantity = df[df['TotalQuantity'] == df['TotalQuantity'].max()]:**

**df['TotalQuantity'].max():** Lấy giá trị lớn nhất trong cột 'TotalQuantity'.

**df['TotalQuantity'] == df['TotalQuantity'].max():** Tạo một mảng boolean, nơi mỗi phần tử là True nếu giá trị 'TotalQuantity' bằng giá trị lớn nhất, ngược lại là False.

**df[df['TotalQuantity'] == df['TotalQuantity'].max()]:** Lọc DataFrame để chỉ giữ lại các dòng có 'TotalQuantity' bằng giá trị lớn nhất. Điều này sẽ là sản phẩm có số lượng bán cao nhất.

**min\_total\_quantity = df[df['TotalQuantity'] == df['TotalQuantity'].min()]:** Tương tự như trên, nhưng lọc để chỉ giữ lại sản phẩm có số lượng bán thấp nhất.

In ra thông tin sản phẩm:

Sử dụng print để hiển thị thông tin của sản phẩm có số lượng bán cao nhất và thấp nhất. Nếu có cột ProductId, thông tin về sản phẩm sẽ được hiển thị để cung cấp cái nhìn chi tiết hơn.

Output:

Ảnh có chứa văn bản, ảnh chụp màn hình, thực đơn

Mô tả được tạo tự động

Hình 3. 15: Kết quả sản phẩm bán chạy nhất và thấp nhất

Tiếp theo em sẽ lọc xem những khách hàng nào là thường xuyên mua hàng nhất và những khách hàng nào ít mua nhất

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Phần mềm đa phương tiện

Mô tả được tạo tự động

Hình 3. 16: Lọc khách hàng

Tiếp theo em sẽ tiến hành trực quan hoá dữ liệu theo tháng và để làm được cái này em sẽ phải Chia cột 'CreatedDate' thành ngày và thời gian

Ảnh có chứa văn bản, phần mềm, Phần mềm đa phương tiện, Hệ điều hành

Mô tả được tạo tự động

Hình 3. 17: Chia cột thành ngày và thời gian

**Chia cột 'CreatedDate' thành ngày và thời gian:** Sử dụng str.split(" ", n=1, expand=True) để tách 'CreatedDate' thành mảng gồm ngày và thời gian, với n=1 chỉ tách một lần.Lưu kết quả vào DataFrame mới date\_time\_parts.

**Tạo cột 'Date' từ phần ngày:** Tạo cột mới 'Date' trong DataFrame và gán giá trị từ cột đầu tiên của date\_time\_parts, tức là phần ngày.

**Chuyển cột 'Date' thành kiểu datetime:** Sử dụng pd.to\_datetime() để chuyển cột 'Date' thành kiểu dữ liệu datetime.

**Thêm các cột mới 'Year', 'Month', và 'Day':** Sử dụng dt.year, dt.month, và dt.day để trích xuất thông tin về năm, tháng và ngày từ cột 'Date'.

**Loại bỏ dòng đầu tiên của DataFrame:** Sử dụng df = df.iloc[4:] để loại bỏ bốn dòng đầu tiên của DataFrame, có thể để loại bỏ dòng chứa dữ liệu không hợp lệ.

Output:

**Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, thực đơn

Mô tả được tạo tự động**

Hình 3. 18: Kết quả sau khi chia

Tiếp theo thực hiện phân tích và biểu đồ hóa tổng lượng 'TotalQuantity' theo tháng từ DataFrame.

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Phần mềm đa phương tiện

Mô tả được tạo tự động

Hình 3. 19: Code vẽ biểu đồ tổng quan cột TotalQuantity

Output:

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, Sơ đồ

Mô tả được tạo tự động

Hình 3. 20: Biểu đồ tổng quan cột TotalQuantity

Bước này giúp công ty có cái nhìn trực quan về sự biến động của tổng lượng 'TotalQuantity' qua các tháng. Bằng cách thực hiện phân tích và biểu đồ hóa, công ty có thể đạt được những lợi ích sau:

**Phát hiện Xu Hướng và Biến Động:** Biểu đồ cung cấp cái nhìn trực quan về xu hướng tăng giảm của lượng 'TotalQuantity' theo tháng. Điều này giúp công ty hiểu rõ hơn về sự biến động trong nhu cầu sản phẩm qua thời gian.

**Lập Kế Hoạch Sản Xuất và Tồn Kho:** Dựa vào biểu đồ, công ty có thể lên kế hoạch sản xuất và quản lý tồn kho dựa trên mô hình sự biến động theo mùa, tháng.

**Phân Tích Hiệu Suất Bán Hàng:** Biểu đồ có thể giúp đánh giá hiệu suất bán hàng trong các chu kỳ kinh doanh khác nhau. Công ty có thể tập trung vào những tháng có nhu cầu cao hơn để tối ưu hóa doanh số bán hàng.

**Đưa Ra Quyết Định Chiến Lược:** Thông qua việc hiểu rõ về biến động của 'TotalQuantity', công ty có thể đưa ra quyết định chiến lược về quảng cáo, khuyến mãi, và các chiến lược bán hàng khác để tối ưu hóa lợi nhuận.

**Dự Báo Nhu Cầu Tương Lai:** Dựa trên xu hướng hiện tại, công ty có thể dự báo nhu cầu tương lai và chuẩn bị các chiến lược sản xuất và tiếp thị.

3.3./ Kết luận chung

**Tối Ưu Hóa Quy Trình Sản Xuất**: Phân tích về "Chemical Using" đã rõ ràng hóa sự chênh lệch giữa trọng lượng đăng ký ('Kg') và trọng lượng thực tế ('RealKg') trong các giao dịch. Điều này quan trọng để điều chỉnh quy trình sản xuất, đảm bảo hiệu suất và giảm thiểu lãng phí nguyên liệu.

**Quản Lý Tồn Kho Hiệu Quả:** Mô hình dự đoán cho "RealKg" mở ra cơ hội để tối ưu hóa quy trình sản xuất và quản lý tồn kho. Điều này giúp giảm tồn kho thừa, phân phối nguồn lực tốt hơn và nâng cao hiệu suất sản xuất tổng thể.

**Chiến Lược Tiếp Thị Có Căn Cứ Thông Tin:** Phân tích bán hàng, đặc biệt là biểu đồ tổng lượng hàng tháng, mang lại cái nhìn sâu sắc về xu hướng và biến động thị trường. Kiến thức này giúp công ty tinh chỉnh chiến lược tiếp thị để đáp ứng một cách linh hoạt với nhu cầu thị trường thay đổi.

3.4./ Ưu điểm và nhược điểm

**Ưu điểm:**

* Hiểu Biết Sâu Sắc với Dữ Liệu: Phân tích cẩn thận của bạn đối với dữ liệu "Chemical Using" và "Sale" cho thấy một hiểu biết sâu sắc về quá trình sản xuất và hiệu suất bán hàng. Việc này làm nền tảng cho quyết định thông tin và chiến lược kinh doanh hiệu quả.
* Mô Hình Dự Đoán Chính Xác: Việc xây dựng mô hình dự đoán cho "RealKg" giúp tối ưu hóa quy trình sản xuất và quản lý tồn kho. Mô hình này cung cấp dự báo chính xác về trọng lượng thực tế, giúp giảm thiểu lãng phí nguyên liệu và tăng cường hiệu suất.
* Khả Năng Tương Tác Với Dữ Liệu: Việc tạo biểu đồ và trực quan hóa dữ liệu giúp tăng cường khả năng tương tác và hiểu biết. Các biểu đồ như biểu đồ cột và biểu đồ tháng cho "TotalQuantity" cung cấp cái nhìn rõ ràng về xu hướng và biến động thị trường.
* Áp Dụng Kiến Thức Trong Thực Tế: Mô hình dự đoán và các biểu đồ giúp áp dụng kiến thức từ dữ liệu vào quyết định kinh doanh hàng ngày. Việc này tạo ra cơ hội để tối ưu hóa chiến lược sản xuất và bán hàng.

**Nhược điểm:**

* Hạn Chế về Số Lượng Dữ Liệu: Nếu có ít dữ liệu, đặc biệt là khi xây dựng mô hình dự đoán, có thể gặp khó khăn trong việc tạo ra mô hình chính xác và độ tin cậy.
* Thiếu Liên Kết Toàn Diện Giữa Các Bộ Dữ Liệu: Việc chưa có sự liên kết chi tiết giữa "Chemical Using" và "Sale" có thể làm giảm khả năng hiểu rõ về cách việc sử dụng hóa chất ảnh hưởng đến doanh số bán hàng.
* Thiếu Phân Tích Chi Tiết về Các Yếu Tố Ảnh Hưởng: Mặc dù có phân tích về 'ScalerId' và 'ChemicalId,' nhưng chưa có sự đào sâu chi tiết về cách chúng ảnh hưởng đến quy trình sản xuất và doanh số bán hàng.
* Cần Thêm Chi Tiết về Chiến Lược Tiếp Thị: Mặc dù có biểu đồ doanh số bán hàng, nhưng có thể cần thêm chi tiết về chiến lược tiếp thị để cung cấp cái nhìn toàn diện về tác động của chiến lược này đối với doanh số bán hàng.

3.5./ Phương hướng phát triển

* Liên Kết Toàn Diện Dữ Liệu: Nâng cao sự liên kết giữa dữ liệu "Chemical Using" và "Sale" để đạt được cái nhìn toàn diện hơn về quá trình sản xuất và doanh số bán hàng. Sự hiểu biết sâu sắc về cách hóa chất ảnh hưởng đến hiệu suất bán hàng có thể dẫn đến các chiến lược tối ưu hóa toàn diện.
* Phân Tích Chi Tiết về Yếu Tố Ảnh Hưởng: Mở rộng phân tích về 'ScalerId' và 'ChemicalId' để hiểu rõ hơn về cách chúng ảnh hưởng đến quy trình sản xuất và doanh số bán hàng. Điều này có thể đưa ra thông tin quý báu về tối ưu hóa nguồn lực và chiến lược tiếp thị.
* Tích Hợp Dữ Liệu Về Chiến Lược Tiếp Thị: Mở rộng phân tích đối với chiến lược tiếp thị để cung cấp cái nhìn chi tiết hơn về tác động của chiến lược này đối với doanh số bán hàng. Điều này giúp tối ưu hóa chiến lược tiếp thị và tăng cường hiệu suất bán hàng.
* Tạo Thêm Mô Hình Dự Đoán: Xem xét việc tạo thêm mô hình dự đoán cho các biến khác ngoài "RealKg" để đánh giá các yếu tố khác ảnh hưởng đến quá trình sản xuất và doanh số bán hàng.
* Kết Hợp Dữ Liệu Với Thông Tin Thị Trường: Tích hợp dữ liệu ngoại vi như thông tin thị trường, xu hướng ngành và sự kiện kinh tế để có cái nhìn toàn diện và chuẩn bị cho các biến động trong môi trường kinh doanh.
* Áp Dụng Kỹ Thuật Học Máy Nâng Cao: Nghiên cứu và áp dụng các kỹ thuật học máy nâng cao để cải thiện độ chính xác của mô hình dự đoán và phân tích. Điều này bao gồm việc sử dụng mô hình phức tạp hơn và tối ưu hóa các siêu tham số.

CHƯƠNG 4./ KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

4.1./ Về bản thân

Trong suốt đợt thực tập tại công ty, em đã thể hiện sự cam kết và sự học hỏi tích cực, đồng thời đưa ra những đóng góp quan trọng cho dự án với đề tài "Chemical Using" và "Sale". Qua đợt thực tập này em thấy bản thân còn rất nhiều thiếu xót, cần học hỏi thêm kinh nghiêm lẫn kiến thức, trao dồi phát triển bản thân để phù hợp yêu cầu về công việc của các doanh nghiệp, công ty, qua đó thấy được môi trường làm việc ở công ty để bản thân thay đổi phù hợp cần phát triển kỹ năng giao tiếp của mình, đặc biệt là trong tình huống khó khăn hoặc xung đột, sẽ giúp em tương tác hiệu quả hơn với đồng nghiệp các anh/chị trong công ty, cơ quan và giải quyết vấn đề một hiệu quả

Lần thực tập này cũng giúp tự đánh giá được điểm mạnh và điểm yếu của bản thận cụ thể là:

**Điểm mạnh:**

* Chăm chỉ và Cam Kết: Em đã thể hiện sự cam kết cao độ và chăm chỉ trong việc thực hiện các nhiệm vụ được giao. Sự sẵn sàng học hỏi và làm việc chăm chỉ là một điểm mạnh quan trọng.
* Kỹ Năng Phân Tích Dữ Liệu: Em đã phát triển kỹ năng phân tích dữ liệu, đặc biệt là khi xử lý và hiểu biết về dữ liệu "Chemical Using" và "Sale". Việc xây dựng mô hình dự đoán cho "RealKg" là một thành tích đáng chú ý.
* Khả Năng Tương Tác với Dữ Liệu: Em đã thể hiện khả năng tương tác tích cực với dữ liệu thông qua việc tạo biểu đồ và trực quan hóa, giúp tạo ra cái nhìn rõ ràng và thuyết phục về thông tin.

**Điểm yếu:**

* Kiến thức còn nhiều hạn chế
* Kỹ năng giao tiếp kém
* Bị động trong công việc

4.2./ Về đề tài

Đề tài “Phân tích dữ liệu” đã mang lại những kiến thức và kỹ năng quan trọng cho bạn. Quá trình phân tích dữ liệu và xây dựng mô hình dự đoán không chỉ giúp bạn hiểu rõ hơn về quá trình sản xuất và bán hàng mà còn cung cấp cơ hội để đề xuất chiến lược cụ thể.

**Về cơ hội và thách thức:**

* Tối Ưu Hóa Quy Trình Sản Xuất: Phân tích dữ liệu "Chemical Using" đã giúp bạn nhận diện cơ hội để tối ưu hóa quy trình sản xuất, đặc biệt là trong việc kiểm soát lượng hóa chất sử dụng và thực hiện theo dung lượng thực tế.
* Chiến Lược Tiếp Thị Hiệu Quả: Phân tích dữ liệu "Sale" đã giúp bạn đánh giá hiệu suất bán hàng, từ đó đưa ra chiến lược tiếp thị hiệu quả hơn, tập trung vào sản phẩm và khách hàng có tiềm năng cao.
* Xây Dựng Mô Hình Dự Đoán: Bạn đã xây dựng một mô hình dự đoán cho biến "RealKg" với sự chênh lệch giữa "Kg" và "RealKg". Điều này giúp dự đoán và điều chỉnh sản xuất theo cách hiệu quả.

**Định Hình Chiến Lược Kinh Doanh:**

* Hiểu Biết Về Khách Hàng: Dữ liệu "Sale" đã cung cấp thông tin về hành vi mua hàng của khách hàng, giúp hiểu rõ hơn về nhu cầu và mong muốn của họ.
* Phát Hiện Xu Hướng Thị Trường: Phân tích thời gian và xu hướng bán hàng giúp bạn dự đoán và phản ánh xu hướng thị trường, từ đó chiếm ưu thế trong quyết định kinh doanh.
* Học Hỏi và Phát Triển Cá Nhân: Phát Triển Kỹ Năng Phân Tích Dữ Liệu: Đề tài đã giúp bạn phát triển và cải thiện kỹ năng phân tích dữ liệu, từ việc tiền xử lý dữ liệu đến xây dựng mô hình dự đoán.
* Giao Tiếp và Trình Bày: Việc trực quan hóa dữ liệu và thuyết trình kết quả phân tích đều thể hiện khả năng giao tiếp và trình bày thông tin một cách rõ ràng và thuyết phục.
* Tích Luỹ Kinh Nghiệm Thực Tế: Đề tài không chỉ là một dự án nghiên cứu, mà còn là một trải nghiệm thực tế quý báu, giúp bạn tích luỹ kinh nghiệm trong ngành công nghiệp.

4.3./ Hướng phát triển cho bản thân

* Liên Tục Học Hỏi: Liên tục cập nhật kiến thức và kỹ năng của mình với các xu hướng mới trong lĩnh vực khoa học dữ liệu và phân tích kinh doanh.
* Tham Gia Dự Án Nâng Cao: Tham gia vào các dự án phức tạp hơn để đối mặt với thách thức và phát triển kỹ năng chuyên sâu hơn.
* Mở Rộng Liên Kết Dữ Liệu: Tiếp tục nghiên cứu và phát triển sự liên kết giữa các yếu tố trong các bộ dữ liệu khác nhau để có cái nhìn toàn diện và chi tiết hơn.
* Chủ Động Hơn Trong Giao Tiếp: Chủ động hơn trong việc chia sẻ ý kiến và đề xuất, đồng thời nâng cao khả năng thuyết trình để truyền đạt thông điệp một cách rõ ràng và thuyết phục.

Trên cùng của Biểu mẫu

TÀI LIỆU THAM KHẢO

**Về Phân Tích Dữ Liệu và Dự Đoán:**

*[1] "Python for Data Science Handbook" của Jake VanderPlas*

* Nguồn tài liệu chi tiết về sử dụng Python trong khoa học dữ liệu và máy học.

*[2] "Introduction to Statistical Learning" của Gareth James, Daniela Witten, Trevor Hastie, và Robert Tibshirani*

* Sách này cung cấp kiến thức căn bản về các mô hình dự đoán và phân tích thống kê sử dụng R.

*[3 ]"Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" của Aurélien Géron*

* Tài liệu hướng dẫn thực hành về machine learning và deep learning sử dụng các thư viện phổ biến.

**Về Trực Quan Hóa Dữ Liệu:**

*[1] "The Visual Display of Quantitative Information" của Edward Tufte*

* Một trong những cuốn sách quan trọng về trực quan hóa dữ liệu.

*[2] "Storytelling with Data" của Cole Nussbaumer Knaflic*

* Hướng dẫn về cách kể chuyện thông qua dữ liệu và biểu đồ.

[Source Code hoàn chỉnh ở đây](https://colab.research.google.com/drive/1a0oruln77bhSv5eyFFP8N8mqdHoyAfXb?usp=sharing)