TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**

A red and blue logo

Description automatically generated

**TẠ HỮU NHÂN – 52000696**

**LÊ MINH QUÂN-52100094**

**TRẦN SỸ THANH TRÂM-52100853**

**BÁO CÁO CUỐI KỲ**

**HỆ THỐNG THƯƠNG MẠI THÔNG MINH**

**Đề tài: Hỗ đề xuất giá bản lẻ cho sản phẩm mới dựa trên dữ liệu về giá trên sàn thương mại điện tử(TMĐT)**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**

A red and blue logo

Description automatically generated

**TẠ HỮU NHÂN – 52000696**

**LÊ MINH QUÂN-52100094**

**TRẦN SỸ THANH TRÂM-52100853**

**BÁO CÁO CUỐI KỲ**

**HỆ THỐNG THƯƠNG MẠI THÔNG MINH**

Người hướng dẫn

**GV. Dương Hữu Phúc**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

Trước tiên, nhóm chúng tôi xin gửi lời cảm ơn chân thành và lòng biết ơn sâu sắc đến thầy **Dương Hữu Phúc** - người đã trực tiếp hướng dẫn, giảng dạy, cung cấp kiến thức cơ sở để cho chúng tôi tôi hoàn thành tốt bài báo cáo cuối kỳ môn “Hệ thống thương mại thông minh”

Tiếp theo, nhóm chúng tôi xin gửi lời cảm ơn chân thành và sâu sắc nhất đến quý thầy cô trong khoa Công nghệ thông tin đã truyền đạt những kiến thức quý báu giúp tôi có thể hoàn thành tốt được bài báo cáo này. Khoa đã luôn sẵn sàng chia sẻ các kiến thức bổ ích cũng như chia sẻ các kinh nghiệm tham khảo tài liệu, giúp ích không chỉ cho việc thực hiện và hoàn thành đề tài nghiên cứu mà còn giúp ích cho việc học tập và rèn luyện trong quá trình thực hành tại trường Đại học Tôn Đức Thắng nói chung.

Cuối cùng, sau khoảng thời gian học tập trên tôi đã hoàn tất đề tài cuối kỳ nhờ vào sự hướng dẫn, giúp đỡ và những kiến thức học hỏi được từ Quý thầy cô. Do giới hạn về mặt kiến thức và khả năng lý luận nên bài báo cáo vẫn còn nhiều thiếu sót và hạn chế, kính mong sự chỉ dẫn và đóng góp của Quý thầy cô giáo để bài Nghiên cứu của nhóm tôi được hoàn thiện hơn. Hơn nữa, nhờ những góp ý từ thầy cô và các bạn, nhóm sẽ hoàn thành tốt hơn ở những bài nghiên cứu trong tương lai. Nhóm chúng tôi mong Quý thầy cô và các bạn– những người luôn quan tâm và hỗ trợ tôi – luôn tràn đầy sức khỏe và sự bình an.

Tôi xin chân thành cảm ơn!

*TP. Hồ Chí Minh, ngày 30. Tháng 11 năm 2024.*

*Tác giả*

*(Ký tên và ghi rõ họ tên)*

**CÔNG TRÌNH ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Nhóm chúng tôi xin cam đoan đây là công trình nghiên cứu của riêng nhóm và được sự hướng dẫn khoa học của GV.Dương Hữu Phúc. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong Dự án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung Dự án của mình**. Trường Đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày 30 tháng 11 năm 2024.*

*Tác giả*

*(Ký tên và ghi rõ họ tên)*

**TÓM TẮT**

Trong bối cảnh cạnh tranh ngày càng gay gắt trên thị trường, việc định giá sản phẩm mới một cách hiệu quả là yếu tố then chốt để đảm bảo thành công kinh doanh. Đề tài này tập trung vào việc ứng dụng các chiến lược phân tích thông minh kinh doanh (Business Intelligence - BI) để định giá sản phẩm mới, dựa trên dữ liệu thực tế thu thập từ các sàn thương mại điện tử.

Dữ liệu được thu thập từ các sàn thương mại điện tử phổ biến, bao gồm thông tin về giá cả, đặc điểm sản phẩm, xu hướng tiêu dùng, và mức độ cạnh tranh trong các phân khúc thị trường cụ thể. Những sản phẩm được chọn lọc để nghiên cứu thuộc các dòng sản phẩm có tiềm năng phát triển, phù hợp với nhu cầu thị trường và xu hướng mua sắm trực tuyến hiện nay.

Quy trình nghiên cứu bao gồm các bước: thu thập và làm sạch dữ liệu, phân tích các yếu tố ảnh hưởng đến giá cả, và áp dụng các công cụ BI để xây dựng mô hình định giá. Các công cụ phân tích như Power BI, Tableau, và các thuật toán học máy (machine learning) được sử dụng để khám phá mối quan hệ giữa giá cả và các yếu tố như thương hiệu, tính năng sản phẩm, và hành vi người tiêu dùng.

Kết quả nghiên cứu cung cấp một hệ thống định giá dự đoán chính xác và linh hoạt, giúp doanh nghiệp tối ưu hóa chiến lược giá cả cho sản phẩm mới. Đồng thời, đề tài cũng đưa ra các khuyến nghị về cách áp dụng chiến lược định giá dựa trên dữ liệu vào thực tế, từ đó nâng cao khả năng cạnh tranh và gia tăng lợi nhuận.

**MỤC LỤC**

DANH MỤC HÌNH ẢNH vi

DANH MỤC BẢNG BIỂU viii

CHƯƠNG 1. GIỚI THIỆU 1

1.1 Bối cảnh 1

1.2 Tuyên bố vấn đề 1

1.3 Mục tiêu 2

1.4 Phạm vi 2

CHƯƠNG 2. TỔNG QUAN TÀI LIỆU 3

2.1 Tổng quan ngành tiêu thụ thực phẩm dạng nước 3

2.2 Các kỹ thuật BI 3

CHƯƠNG 3. PHƯƠNG PHÁP NGHIÊN CỨU 5

3.1 Thu thập dữ liệu 5

3.1.1 Nguồn dữ liệu 5

3.1.2 Làm sạch dữ liệu 9

3.2 Công cụ kỹ thuật 10

3.2.1 Công cụ BI 10

3.2.2 Kỹ thuật phân tích 10

3.3 Phương pháp phân tích 11

CHƯƠNG 4. PHÂN TÍCH DỮ LIỆU 12

4.1 Tìm hiểu hiệu suất bán sản phẩm 12

4.2 Mối quan hệ giữa các biến trong dữ liệu 13

CHƯƠNG 5. ĐỀ XUẤT GIẢI PHÁP 16

5.1 Triển khai giải pháp BI 16

5.2 Kết quả dự kiến 28

CHƯƠNG 6. KẾT LUẬN VÀ ĐỀ XUẤT 29

6.1 Tóm tắt kết quả 29

6.2 Đề xuất mở rộng nghiên cứu 30

TÀI LIỆU THAM KHẢO: 31

# DANH MỤC HÌNH ẢNH

[Hình 1‑1: Mô hình quy trình BI 1](#_Toc184712776)

[Hình 2‑1: khai thác dữ liệu 4](#_Toc184712777)

[Hình 2‑2: mô tả EDA 4](#_Toc184712778)

[Hình 2‑3: quy trình phân tích dự đoán 4](#_Toc184712779)

[Hình 3‑1: file dữ liệu thô sau khi được thu thập từ sàn TMĐT 5](#_Toc184712780)

[Hình 3‑2: dữ liệu thành phần của các loại nước giải khát 5](#_Toc184712781)

[Hình 3‑3: dữ liệu thành phần của các loại sữa bột 6](#_Toc184712782)

[Hình 3‑4: dữ liệu thành phần của các loại sữa chua ăn 6](#_Toc184712783)

[Hình 3‑5: dữ liệu thành phần của các loại sữa chua nước 7](#_Toc184712784)

[Hình 3‑6: dữ liệu thành phần của các loại sữa nước 7](#_Toc184712785)

[Hình 3‑7: file dữ liệu chính sau khi đã làm sạch 9](#_Toc184712786)

[Hình 3‑8: minh họa ứng dụng các kỹ thuật BI trong tableau 11](#_Toc184712787)

[Hình 4‑1 top sản phẩm bán chạy 12](#_Toc184712788)

[Hình 4‑2: top sản phẩm bán tệ nhất 12](#_Toc184712789)

[Hình 4‑3: sơ đồ heatmap giữa các biến trong tập chính 13](#_Toc184712790)

[Hình 4‑4: sơ đồ heatmap giữa thành phần dinh dưỡng NGK với giá thành 14](#_Toc184712791)

[Hình 4‑5: sơ đồ heatmap giữa thành phần dinh dưỡng sữa bột với giá thành 15](#_Toc184712792)

[Hình 4‑6: sơ đồ heatmap giữa thành phần dinh dưỡng SCA với giá thành 15](#_Toc184712793)

[Hình 4‑7: sơ đồ heatmap giữa thành phần dinh dưỡng SCN với giá thành 15](#_Toc184712794)

[Hình 4‑8: sơ đồ heatmap giữa thành phần dinh dưỡng sữa nước với giá thành 15](#_Toc184712795)

[Hình 5‑1: giao diện cho mô hình ML áp dụng cho BI (sữa nước) 18](#_Toc184712796)

[Hình 5‑2: giao diện cho mô hình ML áp dụng cho BI (SCN) 20](#_Toc184712797)

[Hình 5‑3: giao diện cho mô hình ML áp dụng cho BI (SCA) 22](#_Toc184712798)

[Hình 5‑4: giao diện cho mô hình ML áp dụng cho BI (NGK) 23](#_Toc184712799)

[Hình 5‑5: giao diện cho mô hình ML áp dụng cho BI (sữa bột) 26](#_Toc184712800)

[Hình 5‑6: giao diện trả về sản phẩm gần nhất với sản phẩm sắp ra mắt để hỗ trợ việc định giá 27](#_Toc184712801)

[Hình 6‑1: dữ liệu dashboard 29](#_Toc184712802)

[Hình 6‑2: mô hình ML mẫu 29](#_Toc184712803)

# DANH MỤC BẢNG BIỂU

[Bảng 3‑1: mô tả dữ liệu 8](#_Toc184712804)

[Bảng 3‑2: tỉ lệ null của dữ liệu 8](#_Toc184712805)

[Bảng 3‑3: danh sách các file dữ liệu mới 10](#_Toc184712806)

# GIỚI THIỆU

## Bối cảnh

Trong bối cảnh cạnh tranh ngày càng gay gắt trên thị trường, việc định giá sản phẩm mới một cách hiệu quả là yếu tố then chốt để đảm bảo thành công kinh doanh. Đề tài này tập trung vào việc ứng dụng các chiến lược phân tích thông minh kinh doanh (Business Intelligence - BI) để định giá sản phẩm mới, dựa trên dữ liệu thực tế thu thập từ các sàn thương mại điện tử

|  |
| --- |
| BUSINESS INTELLIGENCE (BI) LÀ GÌ? QUY TRÌNH VÀ KINH NGHIỆM TRIỂN KHAI BI  CHO DOANH NGHIỆP | MH - Digital  Hình 1‑1: Mô hình quy trình BI |

## Tuyên bố vấn đề

Dự án này giải quyết vấn đề cốt lõi mà các doanh nghiệp đối mặt khi định giá sản phẩm mới, bao gồm:

* Khó khăn trong việc phân tích và tận dụng lượng lớn dữ liệu từ các sàn thương mại điện tử
* Thiếu mô hình định giá dựa trên dữ liệu đáng tin cậy và phù hợp với từng phân khúc thị trường
* Cạnh tranh về giá với các đối thủ trên sàn thương mại điện tử, ảnh hưởng đến sức hấp dẫn của sản phẩm
* Tối ưu hóa chiến lược giá để vừa đảm bảo lợi nhuận vừa đáp ứng được mong đợi của khách hàng

Những vấn đề này cản trở doanh nghiệp tận dụng tối đa tiềm năng của các sàn thương mại điện tử, gây lãng phí cơ hội và nguồn lực

## Mục tiêu

Dự án hướng đến các mục tiêu cụ thể sau:

* **Tăng doanh thu** thông qua việc xây dựng chiến lược định giá sản phẩm mới dựa trên dữ liệu thực tế từ các sàn thương mại điện tử
* **Cải thiện trải nghiệm khách hàng**, đảm bảo mức giá hợp lý, cạnh tranh và phù hợp với nhu cầu tiêu dùng
* **Tối ưu hóa quy trình định giá** bằng cách áp dụng các công cụ BI và các thuật toán học máy, giúp doanh nghiệp đưa ra quyết định nhanh chóng và hiệu quả hơn
* **Đưa ra các khuyến nghị cụ thể** về chiến lược giá và danh mục sản phẩm để hỗ trợ các hoạt động kinh doanh trong tương lai

## Phạm vi

Dự án tập trung vào phân tích dữ liệu từ các sàn thương mại điện tử lớn tại Việt Nam (như Lazada, và Tiki), trong khoảng thời gian diễn ra môn học. Phạm vi nghiên cứu bao gồm các dòng sản phẩm về sữa và các loại đồ uống phổ biến. Bộ dữ liệu sử dụng trong dự án bao gồm giá bán, mô tả sản phẩm, doanh số bán hàng, và thông tin từ các đối thủ cạnh tranh. Công cụ phân tích chính sẽ là các phần mềm BI Tableau và các thuật toán học máy, nhằm đưa ra kết quả chính xác và ứng dụng thực tiễn cao.

# TỔNG QUAN TÀI LIỆU

## Tổng quan ngành tiêu thụ thực phẩm dạng nước

Ngành thực phẩm dạng nước (như nước uống dinh dưỡng, nước giải khát, và các sản phẩm chức năng) đang phát triển nhanh chóng, thúc đẩy bởi sự gia tăng ý thức về sức khỏe và nhu cầu tiện lợi của người tiêu dùng. Xu hướng hiện tại trong ngành bao gồm:

1. **Tăng trưởng của các sản phẩm dinh dưỡng cá nhân hóa**: Người tiêu dùng ngày càng ưa chuộng các sản phẩm được thiết kế đáp ứng nhu cầu sức khỏe riêng biệt
2. **Ưu tiên sản phẩm bền vững và tự nhiên**: Các thương hiệu chú trọng vào bao bì thân thiện môi trường và thành phần hữu cơ không sử dụng các chất cấm trong công nghiệp hoặc vượt quá mức quy định
3. **Cạnh tranh gay gắt trên giá cả và sự khác biệt hóa sản phẩm**: Các thương hiệu lớn và nhỏ đều tìm cách tối ưu hóa giá trị mang lại cho khách hàng

Ứng dụng BI trong ngành thực phẩm dạng nước cho phép doanh nghiệp phân tích dữ liệu tiêu dùng để nhận diện xu hướng, đánh giá hiệu quả chiến lược giá, và tối ưu hóa danh mục sản phẩm. BI hỗ trợ giải quyết các thách thức như quản lý nhu cầu không đồng đều, tối ưu hóa hàng tồn kho, và cạnh tranh trên thị trường đa dạng.

## Các kỹ thuật BI

Các kỹ thuật chính bao gồm:

* **Khai thác dữ liệu (Data Mining):** Tìm kiếm các mẫu hành vi tiêu dùng từ khối lượng dữ liệu lớn, giúp doanh nghiệp hiểu rõ sở thích và thói quen mua sắm.

|  |
| --- |
| Sự khác nhau giữa Data Mining và Data Science có thể bạn chưa biết  Hình 2‑1: khai thác dữ liệu |

* **Phân tích dữ liệu khám phá (EDA):** tập trung vào việc khám phá, hiểu và tóm tắt dữ liệu nhằm xác định các đặc điểm chính, mối quan hệ, và xu hướng tiềm ẩn trước khi tiến hành các phân tích nâng cao hoặc xây dựng mô hình.

|  |
| --- |
| What is Exploratory Data Analysis?  Hình 2‑2: mô tả EDA |

* **Phân tích dự đoán (Predictive Analytics):** Sử dụng các mô hình thống kê và thuật toán học máy để dự đoán sức tiêu thụ, từ đó xây dựng chiến lược giá.

|  |
| --- |
| Predictive Analytics (Phân tích dự đoán) trong kinh doanh - Minh Duy  Solutions  Hình 2‑3: quy trình phân tích dự đoán |

# PHƯƠNG PHÁP NGHIÊN CỨU

## Thu thập dữ liệu

### Nguồn dữ liệu

Dữ liệu sẽ được thu thập từ sàn thương mại điện tử Lazada bằng selenium. Gồm 2 bộ dữ liệu là dữ liệu bán hàng của sản phẩm (dataset). Với số dòng dữ liệu thô của file dataset.csv ban đầu là 1845 dòng.

|  |
| --- |
| Hình 3‑1: file dữ liệu thô sau khi được thu thập từ sàn TMĐT |

Bộ dữ liệu còn lại là thành phần các chất của sản phẩm phân theo các category:

* Thành phần dinh dưỡng của nước giải khát:

|  |
| --- |
| Hình 3‑2: dữ liệu thành phần của các loại nước giải khát |

* Thành phần dinh dưỡng của sữa bột:

|  |
| --- |
| Hình 3‑3: dữ liệu thành phần của các loại sữa bột |

* Thành phần dinh dưỡng của sữa chua ăn:

|  |
| --- |
| Hình 3‑4: dữ liệu thành phần của các loại sữa chua ăn |

* Thành phần dinh dưỡng của sữa chua nước:

|  |
| --- |
| Hình 3‑5: dữ liệu thành phần của các loại sữa chua nước |

* Thành phần dinh dưỡng của sữa nước:

|  |
| --- |
| Hình 3‑6: dữ liệu thành phần của các loại sữa nước |

Các thuộc tính của dữ liệu chính bao gồm:

Bảng 3‑1: mô tả dữ liệu

|  |  |
| --- | --- |
| **Tên thuộc tính** | **Mô tả** |
| *Product\_name* | Tên của sản phẩm trên các sàn thương mại điện tử |
| *Quantity\_sold* | Số lượng bản ra của sản phẩm đó trên sàn |
| *Brand\_name* | Tên thương hiệu của sản phẩm |
| *Product\_type* | Kiểu đóng gói của sản phẩm khi bán ra |
| *Product\_quantity* | Dung tích của sản phẩm |
| *category* | Danh mục sản phẩm |
| *price* | Giá bán của sản phẩm trên sàn TMĐT |

Tỉ lệ null/khác/none của dữ liệu:

Bảng 3‑2: tỉ lệ null của dữ liệu

|  |  |  |
| --- | --- | --- |
| attribute |  | 0% |
| price |  | 0% |
| quantity\_sold |  | 0% |
| brand | null/\_\_\_ | 0,2% |
| product\_type | khác | 8,6% |
| product\_quantity | null/\_\_\_ | 3,8% |
| average |  | 4,2% |

### Làm sạch dữ liệu

Sau khi rà soát và xác định các vấn đề của dữ liệu thì tiến hành làm sạch lại file dữ liệu:

- Duyệt qua những dòng bị bỏ trống (null) hoặc chứa chuỗi “null” do không thể tìm thấy dữ liệu rồi thực hiện loại bỏ

- Thực hiện loại bỏ những ký tự chuỗi ở cột product\_quantity để chuyển thành dạng số. Sau đó, quy đổi các giá trị về thành “ml”, đối với các giá trị bé hơn 10 ta có thể hiểu đây là gì trị lít hoặc kg nên ta sẽ nhân giá trị đó cho 1000 để đưa về cùng đơn vị

- Kiểm tra lại kiểu dữ liệu của các cột và rà soát lại các dòng một lần nữa để đảm bảo dữ liệu đã được làm sạch

- Tạo file mới chứa dữ liệu đã được làm sạch

- Tạo thêm cột detail\_product để rút gọn tên các sản phẩm, đồng thời ở các bảng thành phần cũng tạo thêm cột tên chứa các dữ liệu giống với cột detail để tạo liên kết

|  |
| --- |
| Hình 3‑7: file dữ liệu chính sau khi đã làm sạch |

* Tổng kết sau khi thực hiện các thao tác lọc sạch và tạo các cột dữ liệu cần thiết ta thu được file dữ liệu chính và thành phần mới:

Bảng 3‑3: danh sách các file dữ liệu mới

|  |  |
| --- | --- |
| **Tên file** | **Số dòng** |
| *Clean\_dataset.csv* | 1326 dòng |
| *DinhDuongNuocGiaiKhat.csv* | 13 dòng |
| *DinhDuongSuaBot.csv* | 18 dòng |
| *DinhDuongSuaChuaAn.csv* | 15 dòng |
| *DinhDuongSuaChuaNuoc.csv* | 18 dòng |
| *DinhDuongSuaNuoc.csv* | 28 dòng |

## Công cụ kỹ thuật

### Công cụ BI

Các công cụ được sữ dụng trong bài báo cáo này gồm:

* **Tableau:** là công cụ BI hàng đầu, cho phép người dùng trực quan hóa dữ liệu, phân tích nhanh chóng, và tạo báo cáo tương tác. Nó hỗ trợ kết nối nhiều nguồn dữ liệu và đưa ra các insight hữu ích để ra quyết định kinh doanh chính xác.
* **Excel:** là công cụ bảng tính phổ biến, hỗ trợ quản lý, tính toán và phân tích dữ liệu hiệu quả. Với các tính năng như hàm tính toán, biểu đồ, và công cụ phân tích dữ liệu, Excel là lựa chọn linh hoạt cho cá nhân và doanh nghiệp trong nhiều lĩnh vực.

### Kỹ thuật phân tích

Bài báo cáo sử dụng các kỹ thuật phân tích EDA để làm rõ mối quan hệ tương quan giữa các biến ảnh hưởng đến giá của sản phẩm. Ngoài ra, các kỹ thuật truy suất, biểu diễn biểu đồ cũng được áp dụng. Cuối cùng, tạo dashboard để giám sát hiệu suất của mô hình.

## Phương pháp phân tích

**Phân Tích Dữ Liệu Khám Phá (EDA)**: Tiến hành phân tích sơ bộ để hiểu dữ liệu và xác định các mẫu hoặc xu hướng.

**Mô Hình Hóa và Mô Phỏng**: Xây dựng các mô hình dự đoán để dự báo doanh thu, tối ưu hóa việc phân bổ sản phẩm, hoặc xác định các yếu tố chính ảnh hưởng đến hiệu suất cửa hàng.

**Trực Quan Hóa**: Tạo các biểu đồ và bảng điều khiển để truyền đạt phát hiện, như bản đồ nhiệt của hiệu suất cửa hàng, biểu đồ xu hướng doanh thu, và dashboard phân khúc khách hàng.

|  |
| --- |
| GitHub - raviatkumar/Hotel-Booking-Analysis-EDA: Using the matplotlib and  seaborn libraries, EDA on Hotel Booking Analysis.  Hình 3‑8: minh họa ứng dụng các kỹ thuật BI trong tableau |

# PHÂN TÍCH DỮ LIỆU

## Tìm hiểu hiệu suất bán sản phẩm

|  |
| --- |
| Hình 4‑1 top sản phẩm bán chạy |

|  |
| --- |
| A graph with different colored bars  Description automatically generated  Hình 4‑2: top sản phẩm bán tệ nhất |

Đánh giá:

* Các sản phẩm bán chạy nhiều nhất thuộc phần nhiều là nhãn hàng Vinamilk. Ngoại trừ các sản phẩm sửa nước ra thì các sản phẩm còn lại khá được ưa chuộng.
* Ngược lại, sản phẩm sữa nước không được ưa chuộng trên các sàn TMĐT. Có thể do độ tuổi quan tâm đến dòng sản phẩm này khó có thể truy cập tới (trẻ em) và phụ huynh thường chọn cách mua sữa tại các siêu thị gần nhà hoặc sử dụng thay thế bằng thực phẩm chức năng nên từ đó mà sản phẩm sửa nước trên các sàn TMĐT ít được chú ý

Kết luận: đối với môi trường TMĐT các mặt hàng thường được giảm giá khi mua với số lượng lớn và các sản phẩm thiết yếu dành cho trẻ em tuổi sơ sinh lại càng được chú tâm. Từ đó, để nâng cao hiệu quả kinh doanh trên môi trường TMĐT ta có thể dựa vào những dữ liệu trên để lựa chọn định hướng kinh doanh cho phù hợp.

## Mối quan hệ giữa các biến trong dữ liệu

Thực hiện phần tích mối quan hệ giữa các biến tong tập dữ liệu chính

|  |
| --- |
| Hình 4‑3: sơ đồ heatmap giữa các biến trong tập chính |

Đánh giá:

1. Hệ số tương quan:

Mỗi ô trong heatmap biểu thị hệ số tương quan giữa hai biến:

Hệ số tương quan dao động từ -1 đến 1:

1 hoặc gần 1: Tương quan dương mạnh (biến này tăng, biến kia cũng tăng).

-1 hoặc gần -1: Tương quan âm mạnh (biến này tăng, biến kia giảm).

0 hoặc gần 0: Không có mối liên hệ rõ ràng.

1. Phân tích tương quan từ heatmap:

quantity\_sold và price: Hệ số tương quan ~0.12, cho thấy mối quan hệ rất yếu giữa số lượng bán được và giá sản phẩm.

quantity\_sold và product\_quantity: Hệ số tương quan ~0.08, gần như không có mối liên hệ đáng kể giữa số lượng bán được và số lượng sản phẩm.

product\_quantity và price: Hệ số tương quan ~0.13, cho thấy mối quan hệ yếu giữa số lượng sản phẩm và giá.

* Kết luận:

Dữ liệu cho thấy các biến được phân tích không có mối quan hệ mạnh với nhau. Điều này có thể ngụ ý rằng các yếu tố như số lượng bán, số lượng sản phẩm, và giá cả bị ảnh hưởng bởi các yếu tố khác ngoài phạm vi của bộ dữ liệu này, hoặc chúng độc lập với nhau. Để đưa ra chiến lược, cần kết hợp thêm các yếu tố khác hoặc áp dụng các phân tích sâu hơn.

Tiếp tục phân tích mối quan hệ giữa giá thành của sản phẩm với các thành phần:

|  |
| --- |
| Hình 4‑4: sơ đồ heatmap giữa thành phần dinh dưỡng NGK với giá thành |

|  |
| --- |
| Hình 4‑5: sơ đồ heatmap giữa thành phần dinh dưỡng sữa bột với giá thành |

|  |
| --- |
| Hình 4‑6: sơ đồ heatmap giữa thành phần dinh dưỡng SCA với giá thành |

|  |
| --- |
| Hình 4‑7: sơ đồ heatmap giữa thành phần dinh dưỡng SCN với giá thành |

|  |
| --- |
| Hình 4‑8: sơ đồ heatmap giữa thành phần dinh dưỡng sữa nước với giá thành |

* Kết luận chung:
* **Tác động của thành phần dinh dưỡng đến giá:**  
  Một số yếu tố như cacao gốc hoặc chất béo có thể có ảnh hưởng lớn đến giá sản phẩm, nhưng nhiều yếu tố khác như Vitamin và khoáng chất có tác động yếu hơn. Nhưng nhìn chung các tác động này là không đáng kể nhưng ta có thể hiểu từ hàm lượng các thành phần cũng như độ đa dạng của thành phần cũng có một phần nào đó làm ảnh hướng đến sự lớn hoặc bé của giá thành hay nói cách khác là chi phí để nguyên liệu để tạo ra sản phẩm sẽ ảnh hương tích cực đến giá thành của sản phẩm
* **Ứng dụng trong chiến lược kinh doanh:**

Tập trung tối ưu hóa các yếu tố có tác động lớn đến giá.

Phát triển các sản phẩm nhắm đến nhóm yếu tố dinh dưỡng có liên quan mật thiết để tăng tính cạnh tranh.

# ĐỀ XUẤT GIẢI PHÁP

## Triển khai giải pháp BI

1. Phát triển Dashboard giám sát hiệu suất cửa hàng theo thời gian thực:

* Tạo ra các bảng điều khiển (dashboard) cho phép giám sát tình trạng bán hàng của các sản phẩm mới so với các sản phẩm cùng loại trên các sàn thương mại điện tử.
* Các chỉ số cần theo dõi bao gồm giá bán, mức độ cạnh tranh, lượng bán, và mức độ hài lòng của khách hàng, giúp đưa ra quyết định giá một cách kịp thời và chính xác.

1. Xây dựng mô hình hổ trợ ra quyết định về giá

* Dựa trên các dữ liệu đã thu thập, thực hiện xây dựng mô hình cung cấp các sản phẩm có chỉ số gần giống nhất với sản phẩm của bản thân.
* Phương pháp so sánh và định giá cho sản phẩm:

Đầu tiên, ta sẽ dựa trên thành phần của sản phẩm được lưu trong các file csv.

File DinhDuongSuaNuoc.csv chứa dữ liệu về thành phần có trong 100ml sữa nước có các thuộc tính:

* Name,
* Energy (kcal),
* Fat (g),
* Protein (g),
* Carbohydrate (g),
* Calcium (mg),
* Zinc (mg),
* Phosporus (mg),
* Selenium (mcg),
* Magnesi (mg),
* Vitamin A (IU),
* Vitamin D3 (IU),
* Vitamin B2 (mcg),
* Natri (mg),
* Vitamin B12 (mcg),
* Vitamin D (IU),
* Vitamin B1 (mcg),
* Fresh Milk Rate,
* Xenlulose (g),
* Vitamin K2 (mcg),
* Vitamin C (mcg),
* GABA (mcg),
* Vitamin K1 (mcg) ,
* Axit Pantothenic (mcg),
* Axit Folic (mcg),
* Biotin (mcg),
* Vitamin PP (mcg),
* DHA (mg),
* Vitamin B6 (mcg),
* Iron (mg),
* Potassium (mg),
* Vitamin B3 (mcg),
* Iot (mcg),
* Lysin (mg),
* Flavor

|  |
| --- |
| **A screenshot of a computer  Description automatically generated**  Hình 5‑1: giao diện cho mô hình ML áp dụng cho BI (sữa nước) |

File DinhDuongSuaChuaNuoc.csv chứa dữ liệu về thành phần có trong 100ml sữa chua nước có các thuộc tính:

* Name,
* Lable,
* Loại Men,
* Vị,
* Năng lượng (kcal),
* Hydrat carbon (g),
* Chất đạm (g),
* Chất béo (g),
* Tỉ lệ sữa tươi,
* Tỉ lệ nha đam,
* Tỉ lệ mứt trái cây,
* Collagen: (mg),
* Tỉ lệ mứt lựu,
* Tỉ lệ thạch dừa,
* Tỉ lệ nếp cẩm,
* Tỉ lệ trân châu đường đen,
* Tỉ lệ mứt dâu,
* Tỉ lệ mứt việt quất,
* Tỉ lệ mứt lá dứa,
* Tỉ lệ sữa bột tách béo,
* Tỉ lệ Sữa chua lên men tự nhiên,
* Calci (mg),
* Natri (mg),
* Lysin (mg),
* Kẽm (mg),
* Magie (mg),
* Axit Folic (µg),
* Vitamin D3 (IU),
* Vitamin A (IU),
* Vitamin B1 (µg),
* Vitamin B6 (µg),
* Vitamin B9 (µg),
* Vitamin B12 (µg),
* Vitamin B3 (µg),
* Vitamin B2 (µg),
* Tỉ lệ ngũ gốc cacao

|  |
| --- |
| **A screenshot of a computer  Description automatically generated**  Hình 5‑2: giao diện cho mô hình ML áp dụng cho BI (SCN) |

File DinhDuongSuaChuaAn.csv chứa dữ liệu về thành phần có trong 100g sữa chua ăn có các thuộc tính:

* Name,
* Lable,
* Loại Men,
* Vị,
* Năng lượng (kcal),
* Hydrat carbon (g),
* Chất đạm (g),
* Chất béo (g),
* Tỉ lệ sữa tươi,
* Tỉ lệ nha đam,
* Tỉ lệ mứt trái cây,
* Collagen: (mg),
* Tỉ lệ mứt lựu,
* Tỉ lệ thạch dừa,
* Tỉ lệ nếp cẩm,
* Tỉ lệ trân châu đường đen,
* Tỉ lệ mứt dâu,
* Tỉ lệ mứt việt quất,
* Tỉ lệ mứt lá dứa,
* Tỉ lệ sữa bột tách béo,
* Tỉ lệ Sữa chua lên men tự nhiên,
* Calci (mg),
* Natri (mg),
* Lysin (mg),
* Kẽm (mg),
* Magie (mg),
* Axit Folic (µg),
* Vitamin D3 (IU),
* Vitamin A (IU),
* Vitamin B1 (µg),
* Vitamin B6 (µg),
* Vitamin B9 (µg),
* Vitamin B12 (µg),
* Vitamin B3 (µg),
* Vitamin B2 (µg),
* Tỉ lệ ngũ gốc cacao

|  |
| --- |
| **A screenshot of a computer  Description automatically generated**  Hình 5‑3: giao diện cho mô hình ML áp dụng cho BI (SCA) |

File DinhDuongNuocGiaiKhat.csv chứa dữ liệu về thành phần có trong 100ml nước giải khát có các thuộc tính:

* Name,
* Vị,
* Năng lượng (kcal),
* Hydrat carbon (g),
* Chất đạm (g),
* Chất béo (g),
* Cynarin (g),
* Vitamin C (mg),
* Natri (mg),
* Tỉ lệ nước ép táo cô đặc,
* Tỉ lệ nước ép lựu cô đặc,
* Tỉ lệ nước ép cam cô đặc,
* Tỉ lệ nước ép,
* Tỉ lệ nước ép nho cô đặc,
* Tỉ lệ nước ép ổi cô đặc,
* Tỉ lệ nước ép kiwi cô đặc,
* Tỉ lệ nước ép đào cô đặc,
* EGCG (mg),
* Polyphenol (mg),
* L-Theanin (mg),
* Vitamin C (g)

|  |
| --- |
| **A screenshot of a computer  Description automatically generated**  Hình 5‑4: giao diện cho mô hình ML áp dụng cho BI (NGK) |

File DinhDuongSuaBot.csv chứa dữ liệu về thành phần có trong 100g sữa bột có các thuộc tính:

* + Name,
  + Lable,
  + Natri (mg),
  + Kali (mg),
  + Clorid (mg),
  + Calci (mg),
  + Phospho (mg),
  + Magnesi (mg),
  + Mangan (μg),
  + Sắt (mg),
  + I-ốt (μg),
  + Kẽm (mg),
  + Đồng (mg),
  + Selen (μg),
  + Vitamin A (IU),
  + Vitamin D3 (IU),
  + Vitamin E (mg α-TE),
  + Vitamin K1 (μg),
  + Vitamin C (mg),
  + Vitamin B1 (mg),
  + Vitamin B2 (mg),
  + Niacin (mg),
  + Vitamin B6 (mg),
  + Acid folic (μg),
  + Acid pantothetic (mg),
  + Vitamin B12 (μg),
  + Biotin (μg),
  + Cholin (mg),
  + Vitamin K2 (μg),
  + Acid linoleic (mg),
  + Acid alpha-linolenic (mg),
  + ARA (Arachidonic acid) (mg),
  + DHA (Docosahexaenoic acid) (mg),
  + Acid Oleic (mg),
  + PUFA (g),
  + MUFA (g),
  + MCT (g),
  + Năng lượng (kcal),
  + Chất đạm (g),
  + Tryptophan (mg),
  + Chất béo (g),
  + Hyđrat cacbon (g),
  + Chất xơ hòa tan (g),
  + GOS (g),
  + 2′-Fucosyllactose (2′-FL) (mg),
  + Taurin (mg),
  + Inositol (mg),
  + L-Carnitin (mg),
  + Nucleotid (mg),
  + Lutein (μg),
  + Độ ẩm (g),
  + Bifidobacterium (cfu),
  + Lactobacillus rhamnosus (cfu),
  + Acid pantothenic (mg),
  + Lysin (mg),
  + Beta-caroten (μg),
  + Collagen (g),
  + Molybden (μg),
  + Vitamin PP (mg),
  + Crôm (μg),
  + Phosphatidylserine (mg),
  + Acid Folic (μg),
  + Acid Pantothenic (μg),
  + Niacin (μg),
  + 2'-Fucosyllactose (2'-FL) (mg),
  + AA (mg),
  + DHA (mg),
  + Vitamin E (IU),
  + CaHMB (g),
  + YBG (mg),
  + Vitamin B9 (mcg),
  + Vitamin B5 (mg),
  + Canxi (mg),
  + Cholesterol (mg),
  + Vitamin B3 (mg)

|  |
| --- |
| **A screenshot of a computer  Description automatically generated**  Hình 5‑5: giao diện cho mô hình ML áp dụng cho BI (sữa bột) |

Giả sử một doanh nghiệp có một sản phẩm mới muốn đưa ra thị trường.Ý tưởng của hệ thống sẽ là dựa vào dữ liệu của thành phần dinh dưỡng, sau đó so sánh với sản phẩm mới của doanh nghiệp, xét xem sản phẩm nào tương đồng với chúng ta về thành phần dinh dưỡng nhất. Ta chỉ cần phân tích giá cả của sản phẩm đấy thì sẽ cho ra được giá bán của sản phẩm doanh nghiệp. Việc phân tích như vậy sẽ giúp ta khoanh vùng dễ hơn, vì hiện nay các sản phẩm trên thị trường có rất nhiều.

Nếu sản phẩm doanh nghiệp của chúng ta tương đồng với một sản phẩm trên thị trường nhưng có chi phí sản xuất thấp hơn, ta có thể đưa sản phẩm ra thị trường với mức giá cạnh tranh. Ngược lại, giá vốn của chúng ta bằng, hoặc quá cao thì ta sẽ cân nhắc hủy dự án.

Đầu vào (sản phẩm của doanh nghiệp chúng ta).Áp dụng khoảng cách Euclide để tìm ra phẩm tương đồng. Đầu ra sẽ là sản phẩm có thành phần dinh dưỡng tương đồng nhất.

Sau khi có sản phẩm có thành phần dinh dưỡng giống chúng ta nhất, hệ thống sẽ hiện bảng giá của các sản phẩm trên thị trường. Ta sẽ dựa vào đó để đưa ra giá bán cuối cùng

|  |
| --- |
| **A screenshot of a computer  Description automatically generated**  Hình 5‑6: giao diện trả về sản phẩm gần nhất với sản phẩm sắp ra mắt để hỗ trợ việc định giá |

## Kết quả dự kiến

Khi các giải pháp BI được triển khai, chúng tôi kỳ vọng đạt được một số kết quả tích cực sau:

1. **Tăng doanh thu**:

Việc định giá chính xác và tối ưu hóa các chiến lược bán hàng dựa trên dữ liệu sẽ giúp tăng cường khả năng cạnh tranh và khuyến khích khách hàng mua nhiều hơn, từ đó tăng doanh thu cho doanh nghiệp.

1. **Cải thiện sự hài lòng của khách hàng**:

Cung cấp giá cả hợp lý và các chương trình ưu đãi sẽ giúp nâng cao trải nghiệm mua sắm của khách hàng. Hơn nữa, việc điều chỉnh giá và sản phẩm dựa trên dữ liệu sẽ giúp phù hợp hơn với nhu cầu và mong muốn của khách hàng.

1. **Giảm chi phí vận hành**:

Việc tối ưu hóa hàng tồn kho và dự báo nhu cầu chính xác sẽ giúp giảm thiểu chi phí vận hành, như chi phí lưu kho và chi phí giao hàng. Đồng thời, việc theo dõi hiệu suất cửa hàng theo thời gian thực sẽ giúp phát hiện sớm các vấn đề và điều chỉnh nhanh chóng, tránh được các chi phí không cần thiết.

Triển khai những giải pháp này sẽ không chỉ giúp doanh nghiệp tăng trưởng bền vững mà còn nâng cao hiệu quả hoạt động và tối ưu hóa chi phí trong quá trình phát triển sản phẩm mới.

# KẾT LUẬN VÀ ĐỀ XUẤT

## Tóm tắt kết quả

|  |
| --- |
| Hình 6‑1: dữ liệu dashboard |

|  |
| --- |
| Hình 6‑2: mô hình ML mẫu |

Từ dữ diệu dashboard và kết quả chạy mô hình ML ta có thể rút ra các kết luận chung về dự án BI:

Ưu điểm: phân tích các dữ liệu có sẵn từ các sàn TMĐT để hỗ trợ việc phân tích chiến lược kinh doanh. Đồng thời mô hình ML giúp gợi ý các sản phẩm tương tự để hỗ trợ việc ra quyết định giúp cải thiện hiệu suất công việc và thời gian cho người kình doanh.

Hạn chế:

* Bị phụ thuộc vào môi trường cung cấp dữ liệu, các dữ liệu tại một môi trường thu tập được cũng chỉ nên áp dụng đối với môi trường đó vì nó sẽ có thể không còn đúng khi áp dụng vào môi trười TMĐT khác hoặc bán sản phẩm trên thị trường thật
* Giá của sản phẩm còn phụ thuộc vào rất nhiều yếu tố khác như giá thành, lợi nhuận mong muốn,… nên việc cung cấp các sản phẩm có giá trị gần giống chỉ mang tính tham khảo.
* Sự hạn chế về dữ liệu cũng là một điểm yếu lớn nếu như mô hình chạy với lượng dữ liệu thu được quá nhỏ sẽ không nói lên được tình hình thực tế của thị trường

## Đề xuất mở rộng nghiên cứu

Với mô hình này có thể áp dụng rộng rãi với các sản phẩm và yêu cầu khác, nhưng phải đảm bảo lượng dữ liệu đủ lớn và thu thập đủ các dữ liệu có tính tất yếu mới có thể áp dụng mô hình.

# TÀI LIỆU THAM KHẢO:

[1]Wikipedia contributors, “Exploratory data analysis,” *Wikipedia*, Nov. 27, 2024. https://en.wikipedia.org/wiki/Exploratory\_data\_analysis

[2]“pandas - Python Data Analysis Library.” https://pandas.pydata.org/

[3]GeeksforGeeks, “Help function in Python,” *GeeksforGeeks*, Jun. 20, 2024. https://www.geeksforgeeks.org/help-function-in-python/

[4]“How can you effectively work with small data sets in Business Intelligence (BI)?,” Dec. 08, 2023. https://www.linkedin.com/advice/3/how-can-you-effectively-work-small-data-sets-mgwqc

[5]“Business Intelligence Topics - Trends and Solutions | Microsoft Power BI.” https://powerbi.microsoft.com/en-us/business-intelligence-topics/