

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ**

**TRUYỀN THÔNG VIỆT – HÀN**

**KHOA KHOA HỌC MÁY TÍNH**



**KHO DỮ LIỆU**

**ĐỀ TÀI: XÂY DỰNG DATAWAREHOUSE BỘ DỮ LIỆU PHÂN TÍCH CẢM XÚC TRONG TIẾNG VIỆT**

**Sinh viên thực hiện: VÕ VĂN NAM 21IT683**

**TRẦN CÔNG HÙNG 21IT615**

**Giảng viên hướng dẫn: THS.MAI LAM**

***Đà Nẵng, tháng 11 năm 2024***

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ**

**TRUYỀN THÔNG VIỆT – HÀN**

**KHOA KHOA HỌC MÁY TÍNH**

A logo of a company

Description automatically generated

**KHO DỮ LIỆU**

**ĐỀ TÀI: XÂY DỰNG DATAWAREHOUSE BỘ DỮ LIỆU PHÂN TÍCH CẢM XÚC TRONG TIẾNG VIỆT**

**Sinh viên thực hiện: VÕ VĂN NAM 21IT683**

**TRẦN CÔNG HÙNG 21IT615**

**Giảng viên hướng dẫn: THS.MAI LAM**

***Đà Nẵng, tháng 11 năm 2024***

**MỞ ĐẦU**

1. **Lý do chọn đề tài**

Trong thời đại số hóa, việc thu thập và phân tích phản hồi của khách hàng đã trở thành yếu tố cốt lõi giúp các doanh nghiệp hiểu rõ hơn về mong muốn và cảm xúc của người dùng. Đặc biệt, phân tích cảm xúc từ văn bản đã trở thành một công cụ hữu ích giúp doanh nghiệp không chỉ theo dõi phản hồi mà còn dự đoán xu hướng thị trường và điều chỉnh chiến lược kinh doanh. Tuy nhiên, đối với tiếng Việt, các giải pháp phân tích cảm xúc vẫn còn nhiều thách thức do ngôn ngữ phức tạp và thiếu tài nguyên dữ liệu phong phú như các ngôn ngữ khác.

Chính vì vậy, đề tài "Thiết kế Data Warehouse bộ dữ liệu phân tích cảm xúc trong văn bản tiếng Việt" được chọn với mục tiêu xây dựng một kho dữ liệu (Data Warehouse) chuyên dụng cho việc lưu trữ, xử lý, và phân tích cảm xúc của người dùng trong các đánh giá tiếng Việt. Kho dữ liệu này không chỉ hỗ trợ việc tổ chức và truy vấn dữ liệu một cách hiệu quả mà còn tạo nền tảng cho các phân tích sâu hơn, góp phần cải thiện trải nghiệm khách hàng. Việc lựa chọn đề tài nhằm đóng góp vào nghiên cứu và phát triển các ứng dụng xử lý ngôn ngữ tự nhiên (NLP) cho tiếng Việt, từ đó hỗ trợ doanh nghiệp nâng cao hiệu quả phân tích và ra quyết định dựa trên phản hồi của người dùng.

1. **Mục tiêu**

Xây dựng kho dữ liệu (Data Warehouse) hỗ trợ lưu trữ, quản lý và phân tích cảm xúc từ các đánh giá bằng tiếng Việt, tạo nền tảng cho các phân tích cảm xúc sâu hơn, từ đó hỗ trợ doanh nghiệp hiểu và cải thiện trải nghiệm khách hàng.

1. **Nội dung nghiên cứu**

Trên cơ sở mục tiêu nghiên cứu, tác giả đã triển khai các nội dung nghiên cứu cụ thể như sau:

* Nghiên cứu các khái niệm cơ bản về kho dữ liệu
* Phân tích dữ liệu Đánh giá phim trên nền tảng IBMD
* Trực quan hóa dữ liệu.
* Nghiên cứu về công cụ và công nghệ hỗ trợ.
* Phân tích các tài liệu và báo cáo về các nghiên cứu tương tự
* Xây dựng mô hình phân tích và trực quan hóa
* Tổng hợp và báo cáo kết quả thực hiện.

1. **Phương pháp nghiên cứu**

* Phương pháp nghiên cứu lý thuyết:
  + Tham khảo các công cụ và nền tảng phân tích dữ liệu hiện có trên thị trường.
  + Nghiên cứu cách thức vận hành của quy trình phân tích và trực quan hóa dữ liệu.
  + Vận dụng các kiến thức đã học để xây dựng quy trình phân tích và trực quan hóa dữ liệu.
* Phương pháp nghiên cứu thực nghiệm
  + Dựa vào các phương pháp nghiên cứu lý thuyết để xây dựng quy trình phân tích và trực quan hóa dữ liệu Marketing, đánh giá và cải tiến.
  + Triển khai thử nghiệm hệ thống kho dữ liệu để kiểm chứng mô hình lý thuyết.

1. **Ý nghĩa khoa học của đề tài**

Đề tài "Thiết kế Data Warehouse bộ dữ liệu phân tích cảm xúc trong văn bản tiếng Việt" mang ý nghĩa khoa học quan trọng trong lĩnh vực xử lý ngôn ngữ tự nhiên (NLP) và quản trị dữ liệu. Cụ thể, đề tài góp phần xây dựng nền tảng dữ liệu chuyên biệt cho phân tích cảm xúc tiếng Việt, một ngôn ngữ phức tạp và ít tài nguyên so với các ngôn ngữ phổ biến khác. Qua việc chuẩn hóa và tổ chức dữ liệu theo mô hình Data Warehouse, nghiên cứu này cung cấp cơ sở hạ tầng dữ liệu đáng tin cậy, giúp các hệ thống phân tích cảm xúc hoạt động hiệu quả và chính xác hơn.

Ngoài ra, đề tài cũng đóng góp vào nghiên cứu ứng dụng Data Warehouse trong các hệ thống NLP, giúp cải thiện khả năng lưu trữ, truy xuất và phân tích dữ liệu lớn. Hệ thống kho dữ liệu này không chỉ hỗ trợ các nghiên cứu phân tích cảm xúc mà còn có thể mở rộng cho các ứng dụng ngôn ngữ tiếng Việt khác, mở ra hướng nghiên cứu và ứng dụng cho các ngôn ngữ địa phương khác có cấu trúc và đặc thù tương tự.

1. **Nội dung thực hiện:**

Chương 1: Nghiên cứu tổng quan

Chương 2: Xây dựng kho dữ liệu (SSIS)

Chương 3: Quá trình SSAS

**LỜI CẢM ƠN**

Đầu tiên, chúng em xin bày tỏ lòng biết ơn sâu sắc đến Ban giám hiệu Trường Đại học Công nghệ thông tin và Truyền thông – Việt Hàn, các thầy cô thuộc khoa Khoa học máy tính đã tận tình giúp đỡ chúng em suốt thời gian học tập và thực hiện đồ án môn Kho dữ liệu.

Đặc biệt, chúng em xin gửi lời cảm ơn chân thành nhất đến TS.Lê Văn Minh và ThS.Mai Lam. Trong quá trình nghiên cứu và thực hiện đồ án, chúng em đã nhận được sự quan tâm giúp đỡ, hướng dẫn chu đáo, tâm huyết của thầy. Thầy đã giúp chúng em tích lũy thêm nhiều kiến thức để có thể cái nhìn sâu sắc và chuyên môn hơn. Đây là hành trang quý báu, là kinh nghiệm để giúp chúng em đi tiếp trên con đường sự nghiệp của mình.

Có lẽ kiến thức là vô hạn mà sự tiếp nhận kiến thức của bản thân mỗi người luôn tồn tại những hạn chế nhất định. Do đó, trong quá trình hoàn thành báo cáo, chắc chắn không tránh khỏi những thiếu sót. Bản thân chúng em rất mong nhận được những góp ý đến từ thầy để báo cáo được hoàn thiện hơn.

Kính chúc thầy sức khỏe, hạnh phúc và thành công trên con đường sự nghiệp giảng dạy.

**Sinh viên**

**NHẬN XÉT**

**(Của giảng viên hướng dẫn)**

Đà Nẵng, tháng 11 năm 2024

**Giảng viên hướng dẫn**

**MỤC LỤC**

[CHƯƠNG 1. GIỚI THIỆU TỔNG QUAN 1](#_Toc182161971)

[1.1 . Phát biểu về dữ liệu 1](#_Toc182161972)

[1.1.1. Mô tả về dữ liệu 1](#_Toc182161973)

[1.1.2. Thuộc tính kho dữ liệu 2](#_Toc182161974)

[1.1.3. Kho dữ liệu đã thực hiện translate, tiền xử lý và trực quan hóa 3](#_Toc182161975)

[CHƯƠNG 2. THIẾT KẾ KHO DỮ LIỆU 5](#_Toc182161976)

[2.1 Thiết kế kho dữ liệu logic 5](#_Toc182161977)

[2.2. Lược đồ Star Schema 6](#_Toc182161978)

[CHƯƠNG 3. CÀI ĐẶT VÀ TRIỂN KHAI KHO DỮ LIỆU 7](#_Toc182161979)

[3.1. Cài đặt kho dữ liệu vật lý 7](#_Toc182161980)

[3.2. Triển khai ETL bằng SSIS 8](#_Toc182161981)

[3.3. Truy vấn dữ liệu 9](#_Toc182161982)

[3.4 Trực quan hóa dữ liệu 10](#_Toc182161983)

[KẾT LUẬN 13](#_Toc182161984)

[1. Những kết quả đạt được 13](#_Toc182161985)

[2. Hạn chế 13](#_Toc182161986)

[3. Hướng phát triển 13](#_Toc182161987)

[4. Hướng nghiên cứu tiếp theo 13](#_Toc182161988)

[TÀI LIỆU THAM KHẢO 14](#_Toc182161989)

**DANH MỤC CÁC BẢNG**

[Bảng 1: Bảng thuộc tính kho dữ liệu 2](#_Toc182161990)

[Bảng 2: Bảng vi\_imdb\_reviews 5](#_Toc182161991)

[Bảng 3: Bảng vi\_imdb\_reviews\_cleaned 5](#_Toc182161992)

[Bảng 4: Bảng fact\_sentiment\_analysis 5](#_Toc182161993)

[Bảng 5: Bảng dim\_sentiment\_category 5](#_Toc182161994)

[Bảng 6: Bảng Dim Review 7](#_Toc182161995)

[Bảng 7: Bảng Fact Sentiment 7](#_Toc182161996)

[Bảng 8: Bảng Dimension Sentiment 7](#_Toc182161997)

**DANH MỤC HÌNH**

[Hình 1: Web chứa dữ liệu gốc 1](#_Toc182162506)

[Hình 2: Dữ liệu chưa được xử lý 3](#_Toc182162507)

[Hình 3: Dữ liệu đã được xử lý 4](#_Toc182162508)

[Hình 4: Quá trình ETL bằng SSIS 8](#_Toc182162509)

[Hình 5: Import dữ liệu từ csv vào DB 8](#_Toc182162510)

[Hình 6: ETL bẳng DIM SENTIMENT 8](#_Toc182162511)

[Hình 7: ETL bảng DIM\_REVIEW 9](#_Toc182162512)

[Hình 8: ELT bảng FACT\_SENTIMENT 9](#_Toc182162513)

[Hình 9: Phân phối số lượng đánh giá 10](#_Toc182162514)

[Hình 10: Word Cloud từ xuất hiện nhiều 11](#_Toc182162515)

[Hình 11: Từ phổ biến trong đánh giá 11](#_Toc182162516)

[Hình 12: Phân phối từ 12](#_Toc182162517)

[Hình 13: Độ dài trung bình của dataset 12](#_Toc182162518)

[Hình 14: Mối tương quan 13](#_Toc182162519)

# CHƯƠNG 1. GIỚI THIỆU TỔNG QUAN

## . Phát biểu về dữ liệu

### Mô tả về dữ liệu

A screenshot of a computer

Description automatically generated

Hình 1: Web chứa dữ liệu gốc

Dữ liệu về các đánh giá phim trên nền tảng IMDB.

Thông qua dữ liệu, người dùng có thể biết được thông tin của **phim**, số tiền đã chi trả cho từng **bộ phim**, phản hồi của khách hàng **về bộ phim đó**,…

Dữ liệu bao gồm **3** cột, **49594** dòng

Link dữ liệu gốc:<https://www.kaggle.com/datasets/namvo1812/vietnamese-sentiment-analysis-50k-imdb?select=VI_IMDB_Clean.csv>

### Thuộc tính kho dữ liệu

Bảng 1: Bảng thuộc tính kho dữ liệu

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Thuộc tính** | **Kiểu dữ liệu** | **Mô tả** |
| 1 | review\_id | INT | ID của đánh giá |
| 2 | raw\_review | TEXT | Là các phản hồi tiếng Anh của người dùng về bộ phim mà họ đã và đang xem chưa được xử lý. |
| 3 | sentiment | VARCHAR(10) | Cảm xúc của đánh giá dựa theo phản hồi(positive, negative) |
| 4 | raw\_vi\_review | TEXT | Là các phản hồi tiếng Việt của người dùng về bộ phim mà họ đã và đang xem chưa được xử lý. |
| 5 | load\_date | DATE | Ngày nạp dữ liệu vào Data Warehouse |
| 6 | token\_count | INT | Số lượng từ trong đánh giá |
| 7 | is\_duplicate | BOOLEAN | Đánh giá có bị trùng lặp hay không |
| 8 | clean\_review | TEXT | Là các phản hồi tiếng Anh của người dùng về bộ phim mà họ đã và đang xem đã được xử lý. |
| 9 | clean\_vi\_review | TEXT | Là các phản hồi tiếng Việt của người dùng về bộ phim mà họ đã và đang xem đã được xử lý. |
| 10 | preprocessed\_on | DATE | Ngày tiền xử lý dữ liệu |
| 11 | fact\_id | INT | ID của bảng fact |
| 12 | word\_count | INT | Tổng số từ trong đánh giá |
| 13 | character\_count | INT | Tổng số ký tự trong đánh giá |
| 14 | created\_date | DATE | Ngày tạo bản ghi |
| 15 | sentiment\_desc | TEXT | Mô tả ngắn gọn về cảm xúc |

A screenshot of a computer

Description automatically generated

Hình 2: Dữ liệu chưa được xử lý

### 1.1.3. Kho dữ liệu đã thực hiện translate, tiền xử lý và trực quan hóa

* Thực hiện các bước tiền xử lý : <https://colab.research.google.com/drive/1BjlTHuUTTNOoIikC1CPiiC8OMWuZeBQN#scrollTo=Yi9P3YW-GjOb>
* Sau khi tiền xử lý, ta được 49594 dòng và 3 cột

A screenshot of a computer

Description automatically generated

Hình 3: Dữ liệu đã được xử lý

# CHƯƠNG 2. THIẾT KẾ KHO DỮ LIỆU

## 2.1 Thiết kế kho dữ liệu logic

* Tầng Staging(Dữ liệu thô)

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Thuộc tính** | **Mô tả** | **Ghi chú** |
| 1 | review\_id | Mã đánh giá | Khóa chính |
| 2 | raw\_review | Đánh giá T.A gốc | Thuộc tính nhận dạng của bảng |
| 3 | raw\_vi\_review | Đánh giá T.V gốc | Thuộc tính nhận dạng của bảng |
| 4 | sentiment | Phân loại cảm xúc | Thuộc tính nhận dạng của bảng |
| 5 | load\_date | Ngày nạp dữ liệu | Thuộc tính nhận dạng của bảng |

Bảng 2: Bảng vi\_imdb\_reviews

* Tầng Processed(Dữ liệu xử lý)

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Thuộc tính** | **Mô tả** | **Ghi chú** |
| 1 | review\_id | Mã đánh giá | Khóa chính |
| 2 | sentiment\_id | Mã cảm xúc | Khóa phụ |
| 3 | clean\_review | Đánh giá T.A đã xử lý | Thuộc tính nhận dạng của bảng |
| 4 | clean\_vi\_review | Đánh giá T.V đã xử lý | Thuộc tính nhận dạng của bảng |
| 5 | sentiment | Phân loại cảm xúc | Thuộc tính nhận dạng của bảng |
| 6 | token\_count | Số lượng từ | Thuộc tính nhận dạng của bảng |
| 7 | is\_duplicate | Lặp từ | Thuộc tính nhận dạng của bảng |
| 8 | preprocessed\_on | Ngày xử lý | Thuộc tính nhận dạng của bảng |

Bảng 3: Bảng vi\_imdb\_reviews\_cleaned

* Tầng Access(Dữ liệu phân tích)

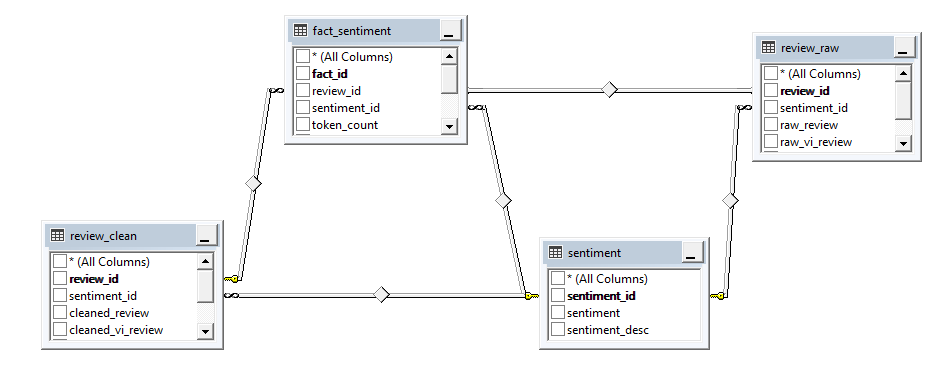
|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Thuộc tính** | **Mô tả** | **Ghi chú** |
| 1 | fact\_id | Mã fact | Khóa chính |
| 2 | review\_id | Mã đánh giá | Khóa phụ |
| 3 | sentiment\_id | Mã cảm xúc | Khóa phụ |
| 4 | sentiment | Phân loại cảm xúc | Thuộc tính nhận dạng của bảng |
| 5 | token\_count | Số lượng từ | Thuộc tính nhận dạng của bảng |
| 6 | word\_count | Tổng số từ | Thuộc tính nhận dạng của bảng |
| 7 | character\_count | Tổng ký tự | Thuộc tính nhận dạng của bảng |
| 8 | sentiment\_embed | Điểm cảm xúc được chuẩn hóa | Thuộc tính nhận dạng của bảng |
| 9 | created\_date | Ngày tạo bảng ghi | Thuộc tính nhận dạng của bảng |

Bảng 4: Bảng fact\_sentiment\_analysis

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Thuộc tính** | **Mô tả** | **Ghi chú** |
| 1 | sentiment\_id | Mã cảm xúc | Khóa chính |
| 2 | sentiment | Phân loại cảm xúc | Thuộc tính nhận dạng của bảng |
| 3 | sentiment\_desc | Mô tả ngắn về cảm xúc | Thuộc tính nhận dạng của bảng |

Bảng 5: Bảng dim\_sentiment\_category

## 2.2. Lược đồ Star Schema



# CHƯƠNG 3. CÀI ĐẶT VÀ TRIỂN KHAI KHO DỮ LIỆU

## 3.1. Cài đặt kho dữ liệu vật lý

* Các bảng:

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Thuộc tính** | **Kiểu dữ liệu** | **Ghi chú** |
| 1 | review\_id | INT | Khóa chính |
| 2 | sentiment\_id | INT | Khóa phụ |
| 3 | clean\_review | TEXT | Thuộc tính nhận dạng của bảng |
| 4 | clean\_vi\_review | TEXT | Thuộc tính nhận dạng của bảng |
| 5 | sentiment | VARCHAT(10) | Thuộc tính nhận dạng của bảng |
| 6 | token\_count | INT | Thuộc tính nhận dạng của bảng |
| 7 | is\_duplicate | BOOLEAN | Thuộc tính nhận dạng của bảng |
| 8 | preprocessed\_on | DATE | Thuộc tính nhận dạng của bảng |

Bảng 6: Bảng Dim Review

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Thuộc tính** | **Mô tả** | **Ghi chú** |
| 1 | fact\_id | Mã fact | Khóa chính |
| 2 | review\_id | Mã đánh giá | Khóa phụ |
| 3 | sentiment\_id | Mã cảm xúc | Khóa phụ |
| 4 | sentiment | Phân loại cảm xúc | Thuộc tính nhận dạng của bảng |
| 5 | token\_count | Số lượng từ | Thuộc tính nhận dạng của bảng |
| 6 | word\_count | Tổng số từ | Thuộc tính nhận dạng của bảng |
| 7 | character\_count | Tổng ký tự | Thuộc tính nhận dạng của bảng |
| 8 | sentiment\_score | Điểm cảm xúc được chuẩn hóa | Thuộc tính nhận dạng của bảng |
| 9 | created\_date | Ngày tạo bảng ghi | Thuộc tính nhận dạng của bảng |

Bảng 7: Bảng Fact Sentiment

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Thuộc tính** | **Mô tả** | **Ghi chú** |
| 1 | sentiment\_id | Mã cảm xúc | Khóa chính |
| 2 | sentiment | Phân loại cảm xúc | Thuộc tính nhận dạng của bảng |
| 3 | sentiment\_desc | Mô tả ngắn về cảm xúc | Thuộc tính nhận dạng của bảng |

Bảng 8: Bảng Dimension Sentiment

## 3.2. Triển khai ETL bằng SSIS

Tổng quan quá trình triển khai ETL bằng SSIS

A screenshot of a computer screen

Description automatically generated

Hình 4: Quá trình ETL bằng SSIS

* Quá trình ETL từ dữ liệu file excel sang Database

A computer screen shot of a computer

Description automatically generated

Hình 5: Import dữ liệu từ csv vào DB

* Quá trình ETL bảng DIM SENTIMENT

A computer screen shot of a computer

Description automatically generated

Hình 6: ETL bẳng DIM SENTIMENT

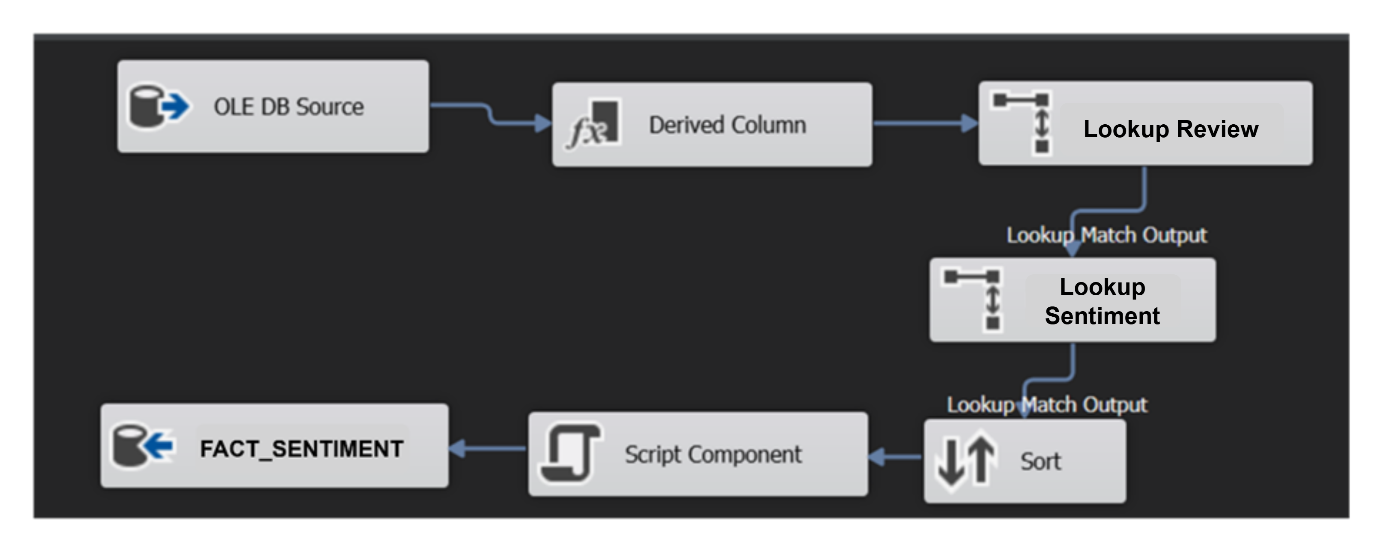
* Quá trình ETL bảng DIM\_REVIEW

A diagram of a process

Description automatically generated

Hình 7: ETL bảng DIM\_REVIEW

* Quá trình ETL bảng FACT\_SENTIMENT



Hình 8: ELT bảng FACT\_SENTIMENT

## 3.3. Truy vấn dữ liệu

* Câu truy vấn 1: Tổng quan về số lượng đánh giá tích cực và tiêu cực:
* Câu lệnh SQL

A screenshot of a computer

Description automatically generated

* Kết quả đầu ra:

A white rectangular object with black text

Description automatically generated with medium confidence

* Câu truy vấn 2: Phân tích các đánh giá có chứa từ khóa nhất định:
* Câu lệnh SQL:

A screenshot of a computer

Description automatically generated

* Kết quả đầu ra:

A screenshot of a computer

Description automatically generated

## 3.4 Trực quan hóa dữ liệu

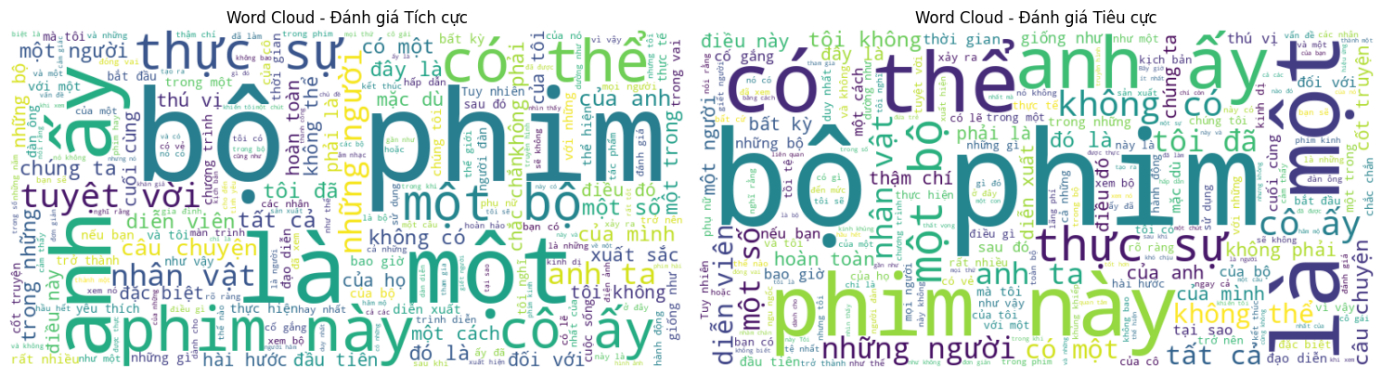
* Phân phối số lượng câu đánh giá trong dataset

A blue and green rectangles

Description automatically generated

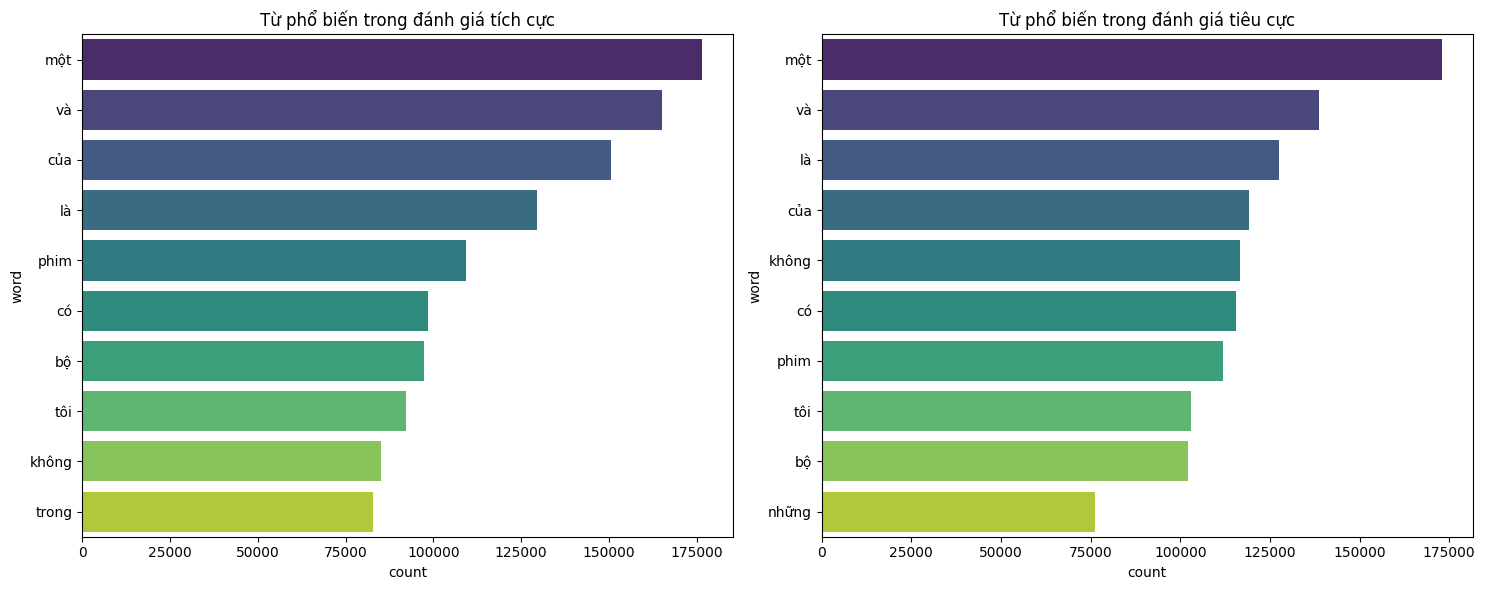
Hình 9: Phân phối số lượng đánh giá

* Word Cloud những từ xuất hiện nhiều trong dataset



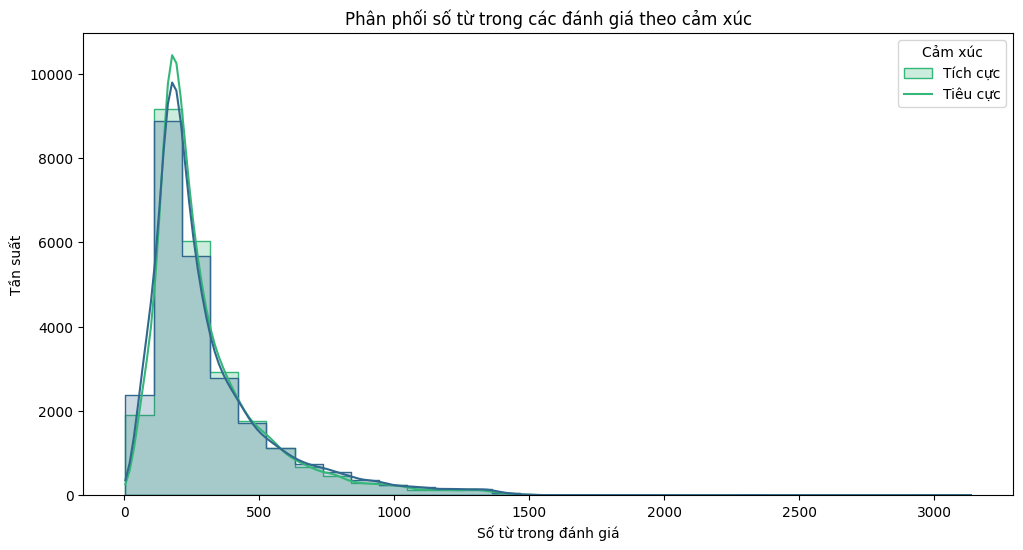
Hình 10: Word Cloud từ xuất hiện nhiều

* Từ phổ biến trong đánh giá tích cực và tiêu cực



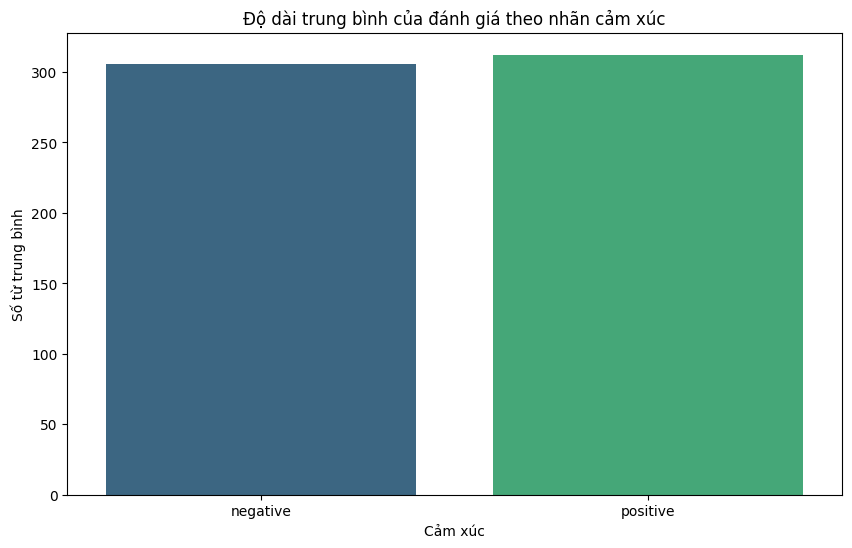
Hình 11: Từ phổ biến trong đánh giá

* Số lượng từ trong các đánh giá theo sentiment



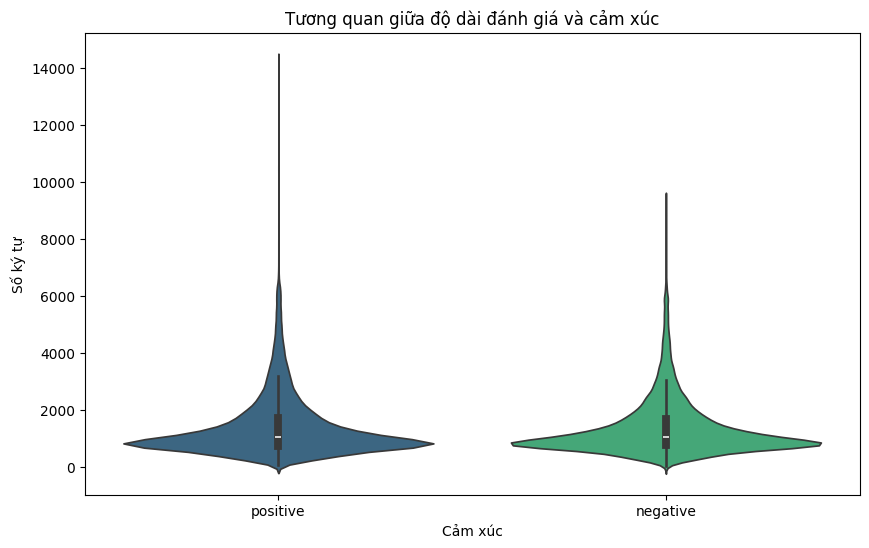
Hình 12: Phân phối từ

* Độ dài trung bình của câu đánh giá trong dataset



Hình 13: Độ dài trung bình của dataset

* Mối tương quan giữa độ dài câu và sentiment



Hình 14: Mối tương quan

# KẾT LUẬN

## 1. Những kết quả đạt được

* Nắm rõ các khái niệm cơ bản về kho dữ liệu, các tính chất của một kho dữ liệu cần có.
* Năm vững kiến thức và có thể vận dụng , xây dựng một kho dữ liệu hoàn chỉnh dùng để khai thác dữ liệu.
* Trang bị kiến thức về các công cụ SSIS, SSAS, SSRS, Data Mining.
* Xây dựng được kho dữ liệu hoàn chỉnh.
* Trình bày tối ưu hóa câu truy vấn.
* Xây dựng được một kho dữ liệu hoàn chỉnh với chủ đề 50.000 câu đánh giá phim từ người dùng của các nhà phát triển trên IMBD.

## 2. Hạn chế

Do thời gian hạn ngắn cộng với khối lượng công việc nhiều nên trong quá trình thực hiện đồ án nhóm còn gặp phải một số vấn đề :

* Chưa áp dụng được so sánh kết quả của nhiều phương pháp Data Mining.
* Quá trình SSIS còn rườm rà chưa được tối ưu.
* Các câu truy vấn MDX chưa nâng cao

## 3. Hướng phát triển

Tiếp tục nghiên cứu và tích hợp các công cụ phân tích dữ liệu nâng cao để tạo báo cáo, thống kê, và bảng điều khiển nhằm tối ưu hóa trong việc thu thập phản hồi người dùng. Mục tiêu là xây dựng một hệ thống báo cáo đa chiều, dễ sử dụng cho các nhà làm phim, đạo diễn,...

Trong tương lai, áp dụng các công nghệ mới, đặc biệt là trí tuệ nhân tạo, để nâng cao khả năng dự đoán xu hướng và hiểu biết sâu hơn về phản hồi khách hàng.

## 4. Hướng nghiên cứu tiếp theo

Nghiên cứu việc tích hợp các công nghệ mới như **Blockchain** để tăng cường tính bảo mật và minh bạch của hệ thống kho dữ liệu Đánh giá cảm xúc. Blockchain có thể đảm bảo tính toàn vẹn của dữ liệu và cung cấp các giải pháp lưu trữ dữ liệu phân tán, giúp ngăn chặn việc thay đổi hoặc giả mạo dữ liệu.

Bên cạnh đó, khám phá các phương pháp mới trong **Phân tích Dữ liệu Lớn** (Big Data Analytics) và **Trực quan hóa Dữ liệu Tương tác** (Interactive Data Visualization) để nâng cao khả năng hiểu biết và ra quyết định dựa trên dữ liệu trong việc thu thập phản hồi. Nghiên cứu ứng dụng **Trí tuệ Nhân tạo** (AI) và **Học máy** (Machine Learning) trong việc phát hiện các mẫu hành vi khách hàng tiềm năng.

Ngoài ra, tiếp tục cải tiến và mở rộng hệ thống để hỗ trợ trong việc thu thập phản hồi trên đa kênh, đa nền tảng như mạng xã hội, email marketing, và các kênh quảng cáo trực tuyến khác, nhằm cung cấp một cái nhìn toàn diện và chính xác hơn về hiệu quả của việc đánh giá cảm xúc của người dùng.

# TÀI LIỆU THAM KHẢO

[1] SQL Server 2012 Tutorials - Analysis Services Multidimensional Modeling

[2] SSRS Tutorials - SQL Server Reporting Service

[3] <https://vi.wikipedia.org/wiki/Microsoft_Visual_Studio>

[4] <https://www.kaggle.com/datasets/namvo1812/vietnamese-sentiment-analysis-50k-imdb?select=VI_IMDB_Clean.csv>

[5] <https://o7planning.org/vi/11385/gioi-thieu-ve-python>

[6] <https://www.youtube.com/watch?v=gH2VXj_-WLc&list=PL2vTwB9LeGfAPsjoD81rX7MP6Zk8iEsY7>

[7] <https://timoday.edu.vn/bai-1-tong-quan-ve-sql-server/>

[8] Slide bài giảng, <https://elib.vku.udn.vn/handle/123456789/2463>

[9] <https://www.youtube.com/watch?v=FQbrUfdre-o&t=41s>

[10] <https://howkteam.vn/course/su-dung-sql-server/khoi-tao-xoa-sua-table-trong-sql-server-1364>

[11] Colab process and visualize dataset, <https://colab.research.google.com/drive/1BjlTHuUTTNOoIikC1CPiiC8OMWuZeBQN#scrollTo=o27zgvS4HO1g>