

**ĐẠI HỌC ĐÀ NẴNG**  
**TRƯỜNG ĐẠI HỌC KINH TẾ**  
**KHOA THỐNG KÊ – TIN HỌC**



# **BÁO CÁO**

## **CHUYÊN ĐỀ TỐT NGHIỆP**

**ĐỀ TÀI**

**ỨNG DỤNG BUSINESS INTELLIGENCE (BI)**  
**PHÂN TÍCH MỨC ĐỘ HÀI LÒNG VỀ SỬ DỤNG DỊCH**  
**VỤ ĐIỆN CỦA KHÁCH HÀNG TẠI CÔNG TY TNHH**  
**MTV ĐIỆN LỰC ĐÀ NẴNG**

<b>Sinh viên thực hiện</b>	<b>: Huỳnh Nguyên Hoàng Trúc</b>
<b>Lớp</b>	<b>: 41K14</b>
<b>Giáo viên hướng dẫn</b>	<b>: ThS. Cao Thị Nhâm</b>

**Đà Nẵng, tháng 12 năm 2019.**

LỜI CẢM ƠN.....	4
DANH MỤC BẢNG.....	5
DANH MỤC HÌNH.....	6
DANH MỤC CÁC KÝ HIỆU, CHỮ VIẾT TẮT.....	9
MỞ ĐẦU.....	10
<b>CHƯƠNG 1 CƠ SỞ LÝ THUYẾT.....</b>	<b>12</b>
1.1. Tổng quan về Business Intelligence (BI).....	12
1.2. Tổng quan về khai phá dữ liệu ( <i>data mining</i> ).....	14
1.3. Quy trình khám phá tri thức.....	15
1.4. Quy trình khai phá dữ liệu.....	16
1.5. Tổng quan về luật kết hợp.....	17
1.5.1 Giới thiệu về luật kết hợp.....	17
1.5.2 Một số tính chất liên quan của luật kết hợp.....	18
1.5.3 Giải thuật Apriori để sinh các luật kết hợp.....	19
<b>CHƯƠNG 2 GIỚI THIỆU CÁC CÔNG CỤ.....</b>	<b>24</b>
2.1. Tổng quan về công cụ trực quan hóa Power BI.....	24
2.2. Tổng quan về Microsoft SQL Server 2014.....	25
2.3. Microsoft SQL Server Analysis Services (SSAS).....	27
2.4. Môi trường triển khai ứng dụng (BIDS).....	28
<b>CHƯƠNG 3 PHÂN TÍCH PHẢN HỒI VỀ VIỆC SỬ DỤNG DỊCH VỤ ĐIỆN TẠI CÔNG TY TNHH MTV ĐIỆN LỰC ĐÀ NẴNG.....</b>	<b>29</b>
3.1. Thực trạng và yêu cầu.....	29
3.2. Đề xuất quy trình.....	29

3.3. Ứng dụng luật kết hợp phân tích phản hồi khách hàng từ 5 điện lực Thành phố Đà Nẵng.....	Error! Bookmark not defined.
<b>CHƯƠNG 4 KẾT QUẢ PHÂN TÍCH.....</b>	<b>48</b>
4.1. Kết quả báo cáo BI.....	48
4.2. Kết quả mô hình luật kết hợp.....	59
<b>KẾT LUẬN.....</b>	<b>1</b>
<b>TÀI LIỆU THAM KHẢO.....</b>	<b>Error! Bookmark not defined.</b>

## **LỜI CAM ĐOAN**

Tôi cam đoan đây là công trình nghiên cứu của riêng tôi. Các số liệu, kết quả nêu trong Báo cáo là hoàn toàn trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

**Sinh viên**

***Huỳnh Nguyên Hoàng Trúc.***

## LỜI CẢM ƠN

Qua thời gian học tập và tích lũy kiến thức dưới sự giảng dạy của các thầy cô tại trường, cùng với 4 tháng thực tập trải nghiệm kinh nghiệm thực tế tại quý công ty TNHH MTV Điện Lực Đà Nẵng. Hôm nay, em đã hoàn thành bài báo cáo thực tập tốt nghiệp với đề tài: ***“Ứng dụng Business Intelligence (BI) phân tích mức độ hài lòng thông qua việc phản hồi về sử dụng điện của khách hàng tại công ty TNHH MTV Điện Lực Đà Nẵng”***.

Lời đầu tiên, em xin chân thành gửi lời cảm ơn tới cô Cao Thị Nhâm – Giảng viên khoa Thống kê – Tin học trường Đại học Kinh tế Đà Nẵng đã tạo điều kiện giúp đỡ, hướng dẫn tận tình để em có thể hoàn thành tốt bài báo cáo này.

Em xin chân thành cảm ơn quý thầy cô khoa Thống Kê – Tin Học cùng các thầy cô trường Đại Học Kinh Tế - Đại Học Đà Nẵng đã tận tình giảng dạy, truyền đạt nhiệt huyết cho em trong suốt bốn năm học qua.

Em xin gửi lời cảm ơn Công ty TNHH MTV Điện lực Đà Nẵng đã nhiệt tình hỗ trợ, hướng dẫn và cung cấp tài liệu thực tế để em có thể hoàn thành báo cáo tốt nghiệp đúng thời gian quy định.

Em xin chân thành cảm ơn!

**Sinh viên**

***Huỳnh Nguyên Hoàng Trúc***

## **DANH MỤC BẢNG**

<b>Bảng 1.1: Thang đo cảm nhận về tính hữu ích.....</b>	<b>22</b>
<b>Bảng 3.1: Mô tả dữ liệu.....</b>	<b>31</b>

## DANH MỤC HÌNH

Hình 1.1. Những thành phần chủ yếu của Business Intelligence (BI).....	13
Hình 1.2. Quy trình khám phá tri thức trong CSDL.....	15
Hình 1.3 Quy trình khai phá dữ liệu.....	16
Hình 1.4 Ví dụ thuật toán Apriori.....	23
Hình 2.1 Trực quan hóa dữ liệu với Power BI.....	24
Hình 2.2 Các môi trường của Power BI.....	24
Hình 2.3 Mô hình kiến trúc SSAS.....	27
Hình 3.1 Đề xuất quy trình khai phá dữ liệu.....	30
Hình 3.2 Dữ liệu thu thập được chưa qua xử lý.....	30
Hình 3.3. Tiền xử lý dữ liệu.....	30
Hình 3.4 Tạo database mới trong Microsoft SQL Server 2014.....	32
Hình 3.5 Import dữ liệu vào database.....	32
Hình 3.6 Import dữ liệu thành công.....	33
Hình 3.7 Sử dụng công cụ Power BI.....	33
Hình 3.8 Tải dữ liệu vào Power BI.....	34
Hình 3.9 Trình điều hướng sau khi tải dữ liệu trong Power BI.....	34
Hình 3.10 Quy trình xây dựng mô hình luật kết hợp.....	34
Hình 3.11 Tạo Data Source Wizard.....	35
Hình 3.12 Tạo Data Source.....	36
Hình 3.13 Tạo Data Source.....	36
Hình 3.14 Tạo Data Source.....	36
Hình 3.15 Tạo Data Source View Wizard.....	37
Hình 3.16 Tạo Data Source View.....	38

Hình 3.17 Quy trình xây dựng mô hình luật kết hợp.....	38
Hình 3.18 Dữ liệu mức độ hài lòng của khách hàng.....	39
Hình 3.19 Giao diện tạo Mining Structures.....	40
Hình 3.20 Tạo Mining Structure.....	40
Hình 3.21 Lựa chọn mô hình luật kết hợp.....	41
Hình 3.22 Chọn thuộc tính để Training.....	41
Hình 3.23 Đặt tên mô hình luật kết hợp.....	42
Hình 3.24 Chọn thuộc tính để Training.....	42
Hình 3.25 Đặt tên mô hình luật kết hợp.....	43
Hình 3.26 Chọn thuộc tính để Training.....	43
Hình 3.27 Đặt tên mô hình luật kết hợp.....	44
Hình 3.28 Chọn thuộc tính để Training.....	44
Hình 3.29 Đặt tên mô hình luật kết hợp.....	45
Hình 3.30 Chọn thuộc tính để Training.....	45
Hình 3.31 Đặt tên mô hình luật kết hợp.....	46
Hình 3.32 Thiết lập tham số mô hình luật kết hợp.....	47
Hình 4.1. Thống kê khách hàng theo từng loại chủ đề yêu cầu.....	48
Hình 4.2 Thống kê khách hàng theo từng loại khách hàng.....	49
Hình 4.3 Thống kê khách hàng theo từng loại khách hàng và chủ đề yêu cầu.....	49
Hình 4.4. Thống kê khách hàng theo từng quận.....	51
Hình 4.5. Thống kê khách hàng theo từng quận và kênh liên lạc.....	52
Hình 4.6. Thống kê khách hàng theo từng quận và loại khách hàng.....	53
Hình 4.7. Thống kê khách hàng theo từng quận và chủ đề yêu cầu.....	55



Hình 4.8. Thống kê mức độ hài lòng khách hàng theo kênh liên lạc.....	57
Hình 4.9. Thống kê khách hàng theo mức độ hài lòng khách hàng và chủ đề yêu cầu.....	57
Hình 4.10. Dashboard thống kê số lượng khách hàng tổng hợp.....	59
Hình 4.11. Kết quả mô hình luật kết hợp.....	59
Hình 4.12. Kết quả mô hình luật kết hợp.....	61
Hình 4.13. Kết quả mô hình luật kết hợp.....	63
Hình 4.14. Kết quả mô hình luật kết hợp.....	65
Hình 4.15. Kết quả mô hình luật kết hợp.....	66
Hình 4.16. Itemsets của mô hình.....	67
Hình 4.17a. Mạng phụ thuộc.....	1
Hình 4.17b. Mạng phụ thuộc.....	1

## DANH MỤC CÁC KÝ HIỆU, CHỮ VIẾT TẮT

Ký hiệu	Ý nghĩa
SQL	Structure Query Language
BI	Business Intelligence
SSAS	SQL Server Analysis Services
BIDS	Business Intelligence Development Studio

# MỞ ĐẦU

## Tính cấp thiết của đề tài

Nền kinh tế nước ta đang ngày một phát triển, kéo theo đó là sự phát triển của các ngành nông nghiệp, công nghiệp, lâm nghiệp, thủy sản. Trong đó, không thể không nhắc đến vai trò quan trọng của ngành Điện lực. Đây chính là một trong những nhân tố thúc đẩy sản xuất, nâng cao năng lực cạnh tranh, thu hút các nguồn vốn đầu tư. Trải qua hơn 60 năm thành lập và phát triển với nhiều dấu mốc đáng nhớ, Điện lực Việt Nam đã tạo nên khối tài sản mang lại giá trị kinh tế xã hội vô giá.

Bên cạnh đó, việc ngành điện ngày một phát triển đồng nghĩa với nhu cầu sử dụng điện của người dân ngày càng tăng cao, kéo theo việc thiếu hụt nguồn điện cung cấp, đặc biệt là vào mùa hè. Từ đó, gây nên các vấn đề như mất điện, mất an toàn điện, cháy nổ trạm điện, thiết lập công tơ... Những việc này đòi hỏi công ty Điện lực phải có những chính sách chăm sóc khách hàng, hỗ trợ xử lý sự cố kịp thời và hợp lý. Để làm được điều đó, chúng ta cần phân tích cụ thể mức độ hài lòng của khách hàng về việc giải đáp thắc mắc cũng như xử lý sự cố điện của Công ty TNHH MTV Điện Lực.

Từ kiến thức của bản thân và những thông tin quý giá từ Công ty TNHH MTV Điện lực Đà Nẵng, tôi mong muốn tiến hành những đánh giá về mức độ hài lòng của khách hàng khi sử dụng dịch vụ chăm sóc khách hàng tại thành phố Đà Nẵng hiện nay bằng cách ứng dụng khai phá dữ liệu để khai thác những thông tin hữu ích từ những kho dữ liệu được tích trữ trong công ty, để từ đó đưa ra một số giải pháp nhằm định hướng việc sử dụng dịch vụ điện cũng như xây dựng chính sách chăm sóc khách hàng hiệu quả hơn trong các hộ gia đình và trong các doanh nghiệp. Do đó tôi đã chọn đề tài “Ứng dụng Business Intelligence phân tích mức độ hài lòng về sử dụng dịch vụ điện của khách hàng tại Công ty TNHH MTV Điện lực Đà Nẵng” làm đề tài báo cáo tốt nghiệp cho mình.

## Mục đích nghiên cứu

Nghiên cứu này được thực hiện với mục tiêu là tìm hiểu về Business Intelligence khai phá dữ liệu luật kết hợp trong khai phá dữ liệu để ứng dụng vào việc phân tích mức độ hài lòng của khách hàng về việc sử dụng dịch vụ điện và xây dựng mô hình. Từ đó đưa ra các giải pháp hỗ trợ các hộ gia đình sử dụng điện, các doanh nghiệp.

### **Kết quả dự kiến**

Phân tích được mức độ hài lòng về việc sử dụng dịch vụ chăm sóc khách hàng ở các điện lực tại thành phố Đà Nẵng.

Tìm ra được luật kết hợp để đánh giá mức độ hài lòng của khách hàng.

### **Phạm vi nghiên cứu**

Nghiên cứu về tình hình phản hồi của khách hàng tại 5 chi nhánh Điện lực: Thanh Khê, Hòa Vang, Cẩm Lệ, Sơn Trà, Hải Châu thuộc Công ty TNHH MTV Điện Lực Đà Nẵng trong năm 2019.

### **Nội dung nghiên cứu**

Nội dung nghiên cứu gồm 3 phần tương ứng với các chương của bài báo cáo này:

Chương 1: Tổng quan.

Chương 2: Giới thiệu các công cụ.

Chương 3: Phân tích mức độ hài lòng thông qua việc phản hồi về sử dụng dịch vụ điện của khách hàng tại Công ty TNHH MTV Điện Lực Đà Nẵng .

Chương 4: Phân tích kết quả .

### **Phương pháp nghiên cứu**

Phương pháp phân tích thông tin thông qua các dữ liệu đã được thu thập tại công ty TNHH MTV Điện Lực Đà Nẵng.

**SỞ LÝ THUYẾT****1.1. Tổng quan về Business Intelligence (BI)*****Business Intelligence (BI)***

Business Intelligence (BI) là quy trình và công nghệ mà các doanh nghiệp dùng để kiểm soát khối lượng dữ liệu khổng lồ, khai phá tri thức giúp cho các doanh nghiệp có thể đưa ra các quyết định hiệu quả hơn trong hoạt động kinh doanh của mình. [1]

Công nghệ BI (BI Technology) cung cấp một cách nhìn toàn cảnh hoạt động của doanh nghiệp từ quá khứ, hiện tại và dự đoán tương lai. Mục đích của BI là hỗ trợ cho doanh nghiệp ra quyết định tốt hơn. Vì vậy một hệ thống BI (BI System) còn được gọi là hệ thống hỗ trợ quyết định (Decision Support System – DSS).

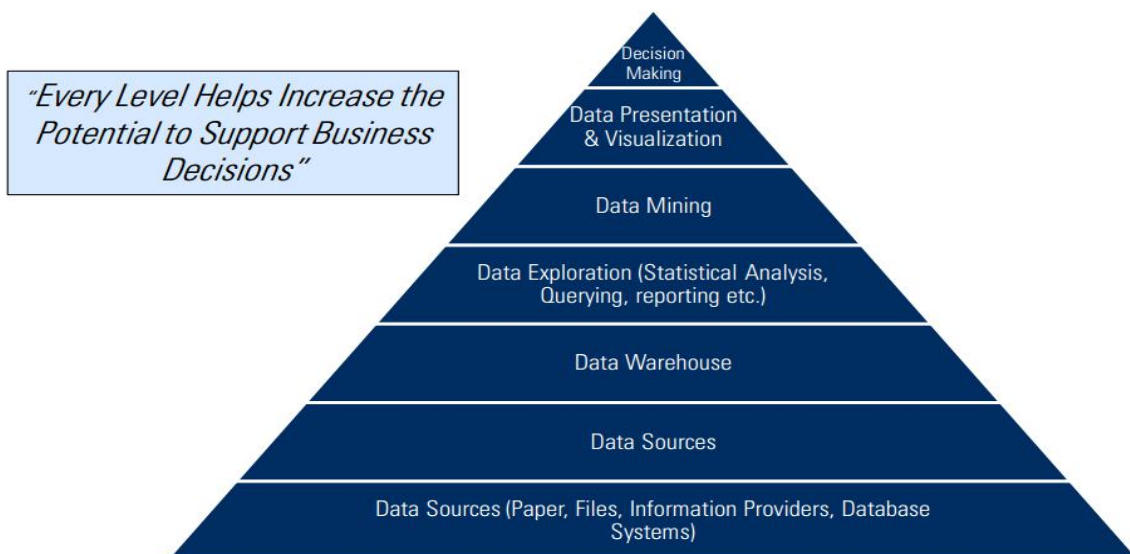
***Lợi ích Business Intelligence (BI)***

- BI làm tăng khả năng kiểm soát thông tin của doanh nghiệp một cách chính xác, hiệu quả. Từ đó có thể phân tích, khai phá tri thức giúp doanh nghiệp có thể dự đoán về xu hướng của giá cả dịch vụ, hành vi khách hàng, phát hiện khách hàng tiềm năng để đề ra những chiến lược kinh doanh phù hợp nhằm tăng khả năng cạnh tranh của doanh nghiệp.

- BI giúp các doanh nghiệp sử dụng thông tin một cách hiệu quả, chính xác để thích ứng với môi trường thay đổi liên tục và cạnh tranh khốc liệt trong kinh doanh.

- Phân tích hành vi khách hàng.
- Đưa ra các chiến lược marketing có hiệu quả.
- Xây dựng chiến lược kinh doanh phù hợp.
- Xác định vị trí và sức cạnh tranh của doanh nghiệp.
- Phát hiện khách hàng tiềm năng của doanh nghiệp.

## Những thành phần chủ yếu của Business Intelligence (BI)



Hình 1.1. Những thành phần chủ yếu của Business Intelligence (BI)

### • Nguồn dữ liệu (Data Sources)

- Là cơ sở dữ liệu thô đến từ nhiều nguồn khác nhau như các ứng dụng business như Human Resource Management (HRM), Customer Relationship Management (CRM), phần mềm bán hàng, website thương mại điện tử... Có thể là bất cứ hệ quản trị cơ sở dữ liệu nào như MySQL, Oracle, MSSQL, DB2... Thường được thiết kế theo mô hình cơ sở dữ liệu quan hệ (vì dạng mô hình này đang rất phổ biến trong thực tế).

- Có thể là bất cứ hệ quản trị cơ sở dữ liệu nào như MySQL, Oracle, MSSQL...

### • Kho dữ liệu (Data Warehouse)

- Là cơ sở dữ liệu được thiết kế theo mô hình khác với cơ sở dữ liệu quan hệ và là nơi lưu trữ dữ liệu lâu dài của tổ chức.

- Dữ liệu của Data Warehouse chỉ có thể đọc, không thể ghi hay update được. Và chỉ được update bởi gói Extract Transform Load chuyển đổi dữ liệu từ Data Sources vào Data Warehouse.

### • Khám phá dữ liệu (Data Exploration)

- Công cụ tích hợp (Integrating Service): Chịu trách nhiệm trung gian vận hành gói ETL để chuyển đổi dữ liệu từ Data Sources vào Data Warehouse.

- Công cụ phân tích (Analysis Service): Chịu trách nhiệm thực thi các Cube được thiết kế dựa trên các Dimension dữ liệu và tri thức nghiệp vụ. Cube chịu trách nhiệm input data từ DW và thực thi theo nghiệp vụ định nghĩa sẵn để trả về output.

- Công cụ báo báo (Reporting Service): Thực thi các report với output nhận được từ Analysis Server. Nơi quản trị tập trung các report trên nền web, các report này có thể được đính kèm vào ứng dụng web hay application.

- **Khai phá dữ liệu (Data Mining)**

- Là quá trình trích xuất thông tin dữ liệu đã qua xử lý (phù hợp với yêu cầu riêng của doanh nghiệp) từ Data Warehouse rồi kết hợp với các thuật toán để đưa ra (hoặc dự đoán) các quyết định có lợi cho việc kinh doanh của doanh nghiệp.

- Đây là một quá trình quan trọng trong BI. Thông thường, nếu một doanh nghiệp muốn sử dụng giải pháp BI thì thường kèm theo về Data Mining.

- **Trình diễn dữ liệu (Data Presentation)**

- Tạo ra các báo cáo, biểu đồ từ quá trình data mining để phục vụ cho nhu cầu của người dùng cuối.

- Ở báo cáo này, chúng ta sẽ tập trung nghiên cứu vào phần trình diễn dữ liệu để khai thác hết tối đa công dụng phục vụ cho công tác trực quan hóa kết quả khảo sát.

## **1.2. Tổng quan về khai phá dữ liệu (*data mining*)**

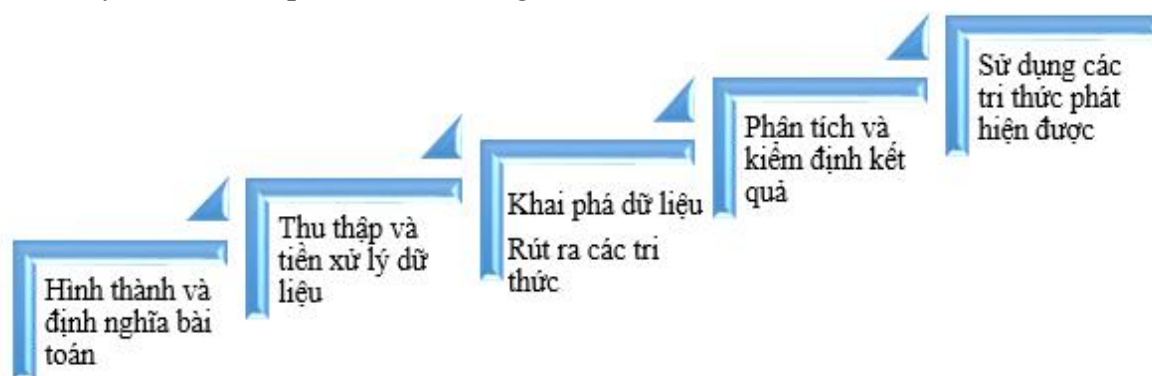
Khai phá dữ liệu được ra đời vào những năm cuối ở thập kỷ 80 của thế kỷ XX. Năm 1989, Fayyad, Piatetsky-Shapiro và Smyth đã dùng khái niệm khám phá tri thức trong cơ sở dữ liệu, để chỉ toàn bộ quá trình phát hiện các tri thức có ích từ các tập dữ liệu lớn. Trong đó khai phá dữ liệu là một bước đặc biệt trong toàn bộ quá trình, sử dụng các giải thuật đặc biệt để chiết xuất ra các mẫu hay các mô hình từ dữ liệu.

Khai phá dữ liệu là quá trình tính toán để tìm ra các mẫu trong các bộ dữ liệu lớn liên quan đến các phương pháp tại giao điểm của máy học, thống kê và các hệ thống cơ sở dữ liệu [2]. Đây là một lĩnh vực liên ngành của khoa học máy tính. Mục tiêu tổng thể của quá trình khai thác dữ liệu là trích xuất thông tin từ một bộ dữ liệu và chuyển nó thành một cấu trúc dễ hiểu để sử dụng tiếp. Ngoài bước phân tích thô, nó còn liên quan tới cơ sở dữ liệu và các khía cạnh quản lý dữ liệu, xử lý dữ liệu trước, suy xét mô hình và suy luận thống kê, các thước đo thú vị, các cân nhắc phức tạp, xuất kết quả về các cấu trúc được phát hiện, hiện hình hóa và cập nhật trực tuyến. Khai thác dữ liệu là bước phân tích của quá trình "khám phá kiến thức trong cơ sở dữ liệu" hoặc KDD.

Khám phá tri thức (KDD) là mục tiêu chính của khai phá dữ liệu, do vậy hai khái niệm đó được xem như hai lĩnh vực tương đương nhau. Nhưng nếu phân chia một cách tách bạch thì khai phá dữ liệu là một bước chính trong quá trình KDD.

### 1.3. Quy trình khám phá tri thức

Quy trình khám phá tri thức bao gồm các bước như sau:



**Hình 1.2. Quy trình khám phá tri thức trong CSDL**

- Bước thứ nhất: Hình thành, xác định và định nghĩa bài toán. Là tìm hiểu lĩnh vực ứng dụng từ đó hình thành bài toán, xác định các nhiệm vụ cần phải hoàn thành. Bước này sẽ quyết định cho việc rút ra được các tri thức hữu ích và cho phép chọn các phương pháp khai phá dữ liệu thích hợp với mục đích ứng dụng và bản chất của dữ liệu.

- Bước thứ hai: Thu thập và tiền xử lý dữ liệu. Là thu thập và tiền xử lý thô, còn được gọi là tiền xử lý dữ liệu nhằm loại bỏ nhiễu, xử lý việc thiếu dữ liệu, biến đổi dữ liệu và rút gọn dữ liệu nếu cần thiết. Bước này thường chiếm nhiều thời gian nhất trong quy trình khám phá tri thức.

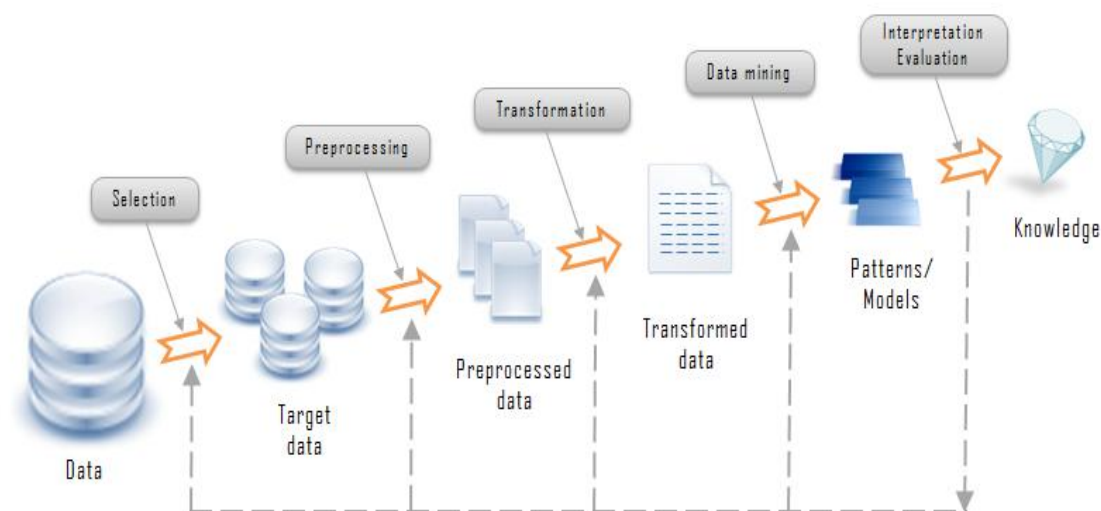
- Bước thứ ba: Khai phá dữ liệu, rút ra các tri thức. Là khai phá dữ liệu hay nói cách khác là trích ra các mẫu và/hoặc các mô hình ẩn dưới các dữ liệu. Giai đoạn này rất quan trọng, bao gồm các công đoạn như: chức năng, nhiệm vụ và mục đích của khai phá dữ liệu, dùng phương pháp khai phá nào?

- Bước thứ tư: Phân tích và kiểm định kết quả: giai đoạn này nhằm làm sáng tỏ các tri thức, đặc biệt là làm sáng tỏ phạm vi của mô tả và dự báo – hai đích chính của hệ thống trong khám phá thực tế. Các thí nghiệm chỉ ra rằng các khuôn mẫu hoặc các mô hình từ dữ liệu không phải luôn luôn được quan tâm hay sử dụng trực tiếp, và quá trình khám phá giá trị tri thức cần phải được lặp đi lặp lại tùy theo đánh giá của tri thức khám phá. Một cách chuẩn để đánh giá kết quả là chia dữ liệu thành hai tập, một tập để huấn luyện và một tập để đánh giá kết quả. Một tập có thể lặp lại quá trình một số lần với sự phân tách khác nhau và kết quả trung bình sẽ dùng để ước lượng luật thực thi.



- Bước thứ năm: Sử dụng các tri thức phát hiện được. Là hiểu tri thức đã tìm được, đặc biệt là làm sáng tỏ các mô tả và dự đoán. Các bước trên có thể lặp đi lặp lại một số lần, kết quả thu được có thể được lấy trung bình trên tất cả các lần thực hiện..

#### 1.4. Quy trình khai phá dữ liệu



**Hình 1.3 Quy trình khai phá dữ liệu.**

- Tạo tập tin dữ liệu đầu vào: Sử dụng Data mining để xác định được những tri thức để chất lọc, từ đó định hướng nhằm tránh tốn thời gian cho những tri thức không cần thiết. Xây dựng tập tin để lưu trữ các dữ liệu đầu vào để máy tính có thể lưu trữ và xử lý.

- Tiền xử lý, làm sạch, mã hóa: Ở bước này cần tiến hành bỏ bớt những dữ liệu rườm rà, không cần thiết, tinh chỉnh lại cấu trúc của dữ liệu và mã hóa chúng để tiện cho quá trình xử lý.

- Rút gọn chiều: Thông thường, một tập dữ liệu có chiều khá lớn sẽ sinh ra một lượng dữ liệu khổng lồ, ví dụ với  $n$  chiều ra sẽ có  $2^n$  tổ hợp. Do đó, đây là bước quan trọng giúp đáng kể hao tổn về tài nguyên trong quá trình xử lý tri thức. Có thể dùng rough set theory (lý thuyết tập thô) để giảm số chiều.

Rough set theory (lý thuyết tập thô) được phát triển trên một nền tảng toán học vững chắc, cung cấp các công cụ hữu ích để giải quyết các bài toán phân tích dữ liệu, phát hiện luật, nhận dạng... với khối lượng dữ liệu lớn, chứa đựng thông tin mơ hồ.

- Khai phá dữ liệu: Lựa chọn thuật toán khai phá dữ liệu và thực hiện khai phá dữ liệu để tìm được các mẫu ý nghĩa, các mẫu này được biểu diễn dưới dạng luật kết hợp, cây quyết định tương ứng với ý nghĩa của nó. Các tác vụ khai phá dữ liệu:

- Đặc trưng (feature)
- Phân biệt (discrimination)
- Kết hợp (association)
- Gom cụm (clusterity)
- Phân tích xu hướng (trend analysis)
- Phân tích độ lệch (deviation analysis)

- Sử dụng các tri thức và khám phá: Qua các tri thức tìm được, cần đánh giá xem những tri thức nào sẽ được sử dụng, những tri thức nào là dư thừa. Biểu diễn tri thức vừa thu nhập được dưới dạng ngôn ngữ tự nhiên và hình thức sao cho người dùng có thể hiểu được những tri thức đó. Tham khảo tiến trình KDD(Knowledge Discovery in Database) để hiểu rõ hơn về khai thác dữ liệu.

## 1.5. Tổng quan về luật kết hợp

### 1.5.1 Giới thiệu về luật kết hợp

Tập toàn bộ các mục  $I = \{i_1, i_2, \dots, i_k\}$ : “tất cả các mặt hàng”. Cho  $D$  là một cơ sở dữ liệu giao dịch như là danh sách các mặt hàng (mục: item) trong một phiếu mua hàng của khách hàng. Giao dịch  $T$  là một tập mục. Một giao dịch  $T$  là một tập con của  $I$ :  $T \subseteq I$ . Mỗi giao dịch  $T$  có một định danh là TID.  $X$  là một tập mục  $X \subseteq I$  và  $T$  là một giao dịch: Gọi  $T$  chứa  $X$  nếu  $X \subseteq T$ . Gọi  $X \Rightarrow Y$  là một “luật kết hợp” nếu  $X \subseteq I$ ,  $Y \subseteq I$  và  $X \cap Y = \emptyset$ .

• Kí hiệu  $\text{support}(X)$  (hoặc  $\text{sup}(X)$ ,  $s(X)$ ) là tỷ lệ phần trăm của các giao dịch hỗ trợ  $X$  trên tổng các giao dịch trong  $D$ , nghĩa là:

$$\text{sup}(X) = \frac{|\{T \in D \mid X \subseteq T\}|}{|D|}$$

Tập mục  $X$  có  $P(X) \geq s > 0$  (với  $s$  cho trước) được gọi là tập phổ biến (frequent set)

Luật kết hợp  $X \Rightarrow Y$  có độ hỗ trợ (support)  $s$  trong CSDL giao dịch  $D$  nếu trong  $D$  có  $s\%$  các giao dịch  $T$  chứa  $X \cup Y$ : chính là xác suất  $P(X \cup Y)$ .  $\text{Support}(X \Rightarrow Y) = P(X \cup Y)$ :  $1 \geq s(X \Rightarrow Y) \geq 0$  (2)

Luật kết hợp  $X \Rightarrow Y$  có độ tin cậy (confidence)  $c$  trong CSDL  $D$  nếu như trong  $D$  có  $c\%$  các giao dịch  $T$  chứa  $X$  thì cũng chứa  $Y$ : chính là xác suất  $P(Y|X)$

$\text{Confidence}(X \Rightarrow Y) = P(Y|X)$ :  $1 \geq c(X \Rightarrow Y) \geq 0$  (3) 6 Luật  $X \Rightarrow Y$  được gọi là đảm bảo độ hỗ trợ  $s$  trong  $D$  nếu  $s(X \Rightarrow Y) \geq s$ .

Luật  $X \Rightarrow Y$  được gọi là đảm bảo độ tin cậy  $c$  trong  $D$  nếu  $c(X \Rightarrow Y) \geq c$ . Luật thỏa mãn cả hai ngưỡng hỗ trợ tối thiểu ( $\min\_sup$ ) và ngưỡng tin cậy tối thiểu ( $\min\_conf$ ) được gọi là luật mạnh. Độ hỗ trợ và độ tin cậy nhận giá trị trong khoảng từ 0% đến 100% hoặc từ 0.0 đến 1.0. giá trị  $\min\_sup$  và  $\min\_conf$  được xác định bởi người dùng hoặc hệ chuyên gia. Như vậy, quá trình khai phá luật kết hợp phải được thực hiện thông qua hai bước:

**Bước 1:** Tìm tất cả các tập mục mà có độ hỗ trợ lớn hơn độ hỗ trợ tối thiểu do người dùng xác định. Các tập mục thỏa mãn độ hỗ trợ tối thiểu được gọi là các tập mục phổ biến.

**Bước 2:** Dùng các tập mục phổ biến để sinh ra các luật mong muốn. Ý tưởng chung là nếu gọi  $ABCD$  và  $AB$  là các tập mục phổ biến, thì chúng ta có thể xác định luật nếu  $AB \Rightarrow CD$  giữ lại với tỷ lệ độ tin cậy

$$conf = \frac{\sup(ABCD)}{\sup(AB)}$$

Nếu  $conf \geq \min\_conf$  thì luật được giữ lại (luật này sẽ thỏa mãn độ hỗ trợ tối thiểu vì  $ABCD$  là phổ biến).

### 1.5.2 Một số tính chất liên quan của luật kết hợp

Với tập mục phổ biến, có 3 tính chất:

- Tính chất 1 (Độ hỗ trợ của tập con): Với  $A$  và  $B$  là tập các mục, nếu  $A \subseteq B$  thì  $\sup(A) \geq \sup(B)$ .

Điều này là rõ ràng vì tất cả các giao tác của  $D$  hỗ trợ  $B$  thì cũng hỗ trợ  $A$ .

- Tính chất 2: Một tập chứa một tập không phổ biến thì cũng là tập không phổ biến.

Nếu một mục trong  $B$  không có độ hỗ trợ tối thiểu trên  $D$  nghĩa là  $\sup(B) < \minsup$  thì một tập con  $A$  của  $B$  sẽ không phải là một tập phổ biến vì  $\sup(A) \leq \sup(B) < \minsup$  (theo tính chất 1).

- Tính chất 3: Các tập con của tập phổ biến cũng là tập phổ biến.

Nếu mục  $B$  là mục phổ biến trên  $D$ , nghĩa là  $\sup(B) \geq \minsup$  thì mọi tập con  $A$  của  $B$  là tập phổ biến trên  $D$  vì  $\sup(A) \geq \sup(B) \geq \minsup$ .

Với luật kết hợp, có 4 tính chất:

- Tính chất 1: Không hợp các luật kết hợp

Nếu có  $X \Rightarrow Z$  và  $Y \Rightarrow Z$  trong  $D$  thì không nhất thiết  $X \cup Y \Rightarrow Z$  là đúng

Xét trường hợp  $X \cap Z = \emptyset$  và các tác vụ trong D hỗ trợ Z nếu và chỉ nếu chúng hỗ trợ mỗi X hoặc Y, khi đó luật  $X \cup Y \Rightarrow Z$  có độ hỗ trợ 0%.

Tương tự:  $X \Rightarrow Y \wedge X \Rightarrow Z$  suy ra  $X \Rightarrow Y \cup Z$

- Tính chất 2: Không tách luật

Nếu  $X \cup Y \Rightarrow Z$  thì  $X \Rightarrow Z$  và  $Y \Rightarrow Z$  chưa chắc xảy ra. Ví dụ trường hợp Z có mặt trong một giao tác chỉ khi cả hai X và Y cũng có mặt, tức là  $\text{sup}(X \cup Y) = \text{sup}(Z)$ , nếu độ hỗ trợ của X và Y đủ lớn hơn  $\text{sup}(X \cup Y)$ , tức là  $\text{sup}(X) > \text{sup}(X \cup Y)$  và  $\text{sup}(Y) > \text{sup}(X \cup Y)$  thì hai luật riêng biệt sẽ không đủ độ tin cậy. Tuy nhiên đảo lại:  $X \Rightarrow Y \cup Z$  suy ra  $X \Rightarrow Y \wedge X \Rightarrow Z$ .

- Tính chất 3: Các luật kết hợp không có tính bắc cầu

Nếu  $X \Rightarrow Y$  và  $Y \Rightarrow Z$ , chúng ta không thể suy ra  $X \Rightarrow Z$ .

Ví dụ: Giả sử  $T(X) \subset T(Y) \subset T(Z)$ , ở đó  $T(X)$ ,  $T(Y)$ ,  $T(Z)$  tương ứng là các giao dịch chứa X, Y, Z, và độ tin cậy cực tiểu minconf.

$\text{conf}(X \Rightarrow Y) = \text{conf}(Y \Rightarrow Z) = \text{minconf}$  thế thì:  $\text{conf}(X \Rightarrow Z) = \text{minconf}^2 < \text{minconf}$  vì  $\text{minconf} < 1$ , do đó luật  $X \Rightarrow Z$  không đủ độ tin cậy.

- Tính chất 4: Nếu  $A \Rightarrow (L - A)$  không thỏa mãn độ tin cậy cực tiểu thì luật  $B \Rightarrow (L - B)$  cũng không thỏa mãn, với các tập mục L, A, B và  $B \subseteq A \subset L$

Vì  $\text{supp}(B) \geq \text{supp}(A)$  (theo tính chất 1) và định nghĩa độ tin cậy, chúng ta nhận được:

$$\text{conf}(B \Rightarrow (L - B)) = \frac{\text{supp}(L)}{\text{supp}(B)} \leq \frac{\text{supp}(L)}{\text{supp}(A)} < \text{minconf}$$

Cũng như vậy: Nếu có  $(L - C) \Rightarrow C$  thì ta cũng có luật  $(L - D) \Rightarrow D$ , với  $D \subseteq C$  và  $D \neq \emptyset$ .

Bởi vì  $D \subseteq C$  nên  $(L - D) \supseteq (L - C)$ , do đó  $\text{supp}(L - D) \leq \text{supp}(L - C)$

$$\longrightarrow \frac{\text{supp}(L)}{\text{supp}(L - D)} \geq \frac{\text{supp}(L)}{\text{supp}(L - C)} \geq \text{minconf}$$

### 1.5.3 Giải thuật Apriori để sinh các luật kết hợp

Để hình dung rõ hơn các giai đoạn trong quá trình phát hiện luật kết hợp, chúng ta sẽ xem xét giải thuật Apriori phát hiện luật kết hợp từ cơ sở dữ liệu tác vụ. [3]

#### - Tính chất Apriori

Mọi tập con khác rỗng của tập mục phổ biến cũng phải phổ biến. Ví dụ, nếu {bia, bìm, hạnh nhân} là phổ biến thì {bia, bìm} cũng vậy, mọi giao dịch chứa {bia, bìm,

hạnh nhân} cũng chứa {bia, bím}.

### - Giải thuật Apriori

Duyệt (Scan) toàn bộ cơ sở dữ liệu giao dịch để có được độ hỗ trợ (support) S của 1- itemset, so sánh S với độ hỗ trợ tối thiểu min\_sup, để có được 1-itemset (F1)

Sử dụng Fk-1 nối (kết nối) Fk-1 để sinh ra candidate k-itemset. Loại bỏ các itemsets không phải là tập phổ biến thu được k-itemset ( bước cắt tỉa)

Tiếp tục duyệt cơ sở dữ liệu giao dịch để có được độ hỗ trợ S của mỗi candidate k-itemset, so sánh S với min\_sup để thu được tập phổ biến k-itemset (Fk)

Lặp lại từ bước 2 cho đến khi Candidate set (C) trống (không tìm thấy tập phổ biến)

Với mỗi tập phổ biến I, sinh tất cả các tập con s không rỗng của I

Với mỗi tập con s không rỗng của I, sinh ra các luật  $s \Rightarrow (I-s)$  nếu độ tin cậy (Confidence) của nó  $\geq \text{min\_conf}$

### - Thuật toán Apriori

#### Các kí hiệu:

- $L_k$ : Tập các k-mục phổ biến (large k-itemset) (tức tập các itemset có sup tối thiểu và có lực lượng bằng k).

Mỗi phần tử của tập này có 2 trường: itemset và support-count.

- $C_k$ : Tập các candidate k-itemset (tập các tập k-mục ứng cử viên). Mỗi phần tử trong tập này cũng có 2 trường itemset và support-count.

Nội dung thuật toán Apriori được trình bày như sau:

**Input:** Tập các giao dịch D, ngưỡng support tối thiểu minsup

**Output:** L- tập mục phổ biến trong D

#### Method:

$L_1 = \{\text{large 1-itemset (tập 1- mục phổ biến)}\}$  //tìm tất cả các tập mục phổ biến:  
nhận được  $L_1$

**For** ( $k=2$ ;  $L_{k-1} \neq \emptyset$ ;  $k++$ ) **do**

#### Begin

$C_k = \text{apriori-gen}(L_{k-1})$ ; //sinh ra tập ứng cử viên từ  $L_{k-1}$

**For** (mỗi một giao dịch  $T \in D$ ) **do**

#### Begin

$C_T = \text{subset}(C_k, T)$ ; //lấy tập con của T là ứng cử viên trong  $C_k$

**For** (mỗi một ứng cử viên  $c \in C_T$ ) **do**

$c.\text{count}++$ ; //tăng bộ đếm tần xuất 1 đơn vị

**End;**

$L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\}$

**End;**

return  $\cup_k L_k$

Trong thuật toán này, giai đoạn đầu đơn giản chỉ là việc đếm support cho các mục(item). Để xác định tập 1-mục phổ biến ( $L_1$ ), người ta chỉ giữ lại các mục (item) mà sup của nó lớn hơn hoặc bằng *minsup*.

Tiếp theo CSDL D sẽ được quét để tính support cho mỗi ứng viên trong  $C_k$ . Để việc đếm được nhanh, cần phải có một giải pháp hiệu quả để xác định các ứng viên trong  $C_k$  là có mặt trong một giao dịch T cho trước.

Vấn đề sinh tập candidate (tập ứng viên) của Apriori – Hàm Apriori\_gen: Hàm Apriori\_gen với đối số là  $L_{k-1}$  (tập các large(k-1)-itemset) sẽ cho lại kết quả là một superset, tập của tất cả các large k – itemset. Sơ đồ sau là thuật toán cho hàm này.

**Input:** tập mục phổ biến  $L_{k-1}$  có kích thước k-1

**Output:** tập ứng cử viên  $C_k$

**Method:**

**function** apriori-gen( $L_{k-1}$ : tập mục phổ biến có kích thước k-1)

**Begin**

**For** (mỗi  $L_1 \in L_{k-1}$ ) **do**

**For** (mỗi  $L_2 \in L_{k-1}$ ) **do**

**begin**

**If**  $((L_1[1]=L_2[1]) \cap (L_1[2]=L_2[2]) \cap \dots \cap (L_1[k-2]=L_2[k-2]) \cap (L_1[k-1]=L_2[k-1]))$

**then**

$c = L_1 \oplus L_2$ ; // kết nối  $L_1$  với  $L_2$  sinh ra ứng cử viên C

**If** has\_infrequent\_subset( $c, L_{k-1}$ ) **then** //có tập con không phổ biến của c trong  $L_{k-1}$

remove (C) // bước tĩa (xoá ứng cử viên C)

**else**  $C_k = C_k \cup \{c\}$ ; kết tập c vào  $C_k$

**end;**

Return  $C_k$ ;

**End;**

Với nội dung trên, ta thấy hàm này có 2 bước:

- Bước nối (join step)
- Bước cắt tỉa (prune step)

Hàm Subset (trong bước tỉa) Các tập ứng cử viên  $C_k$  được lưu trữ trong một cây băm.

- Nút lá của cây băm chứa danh sách một tập mục và đếm
- Các nút trong chứa ở trong bảng băm

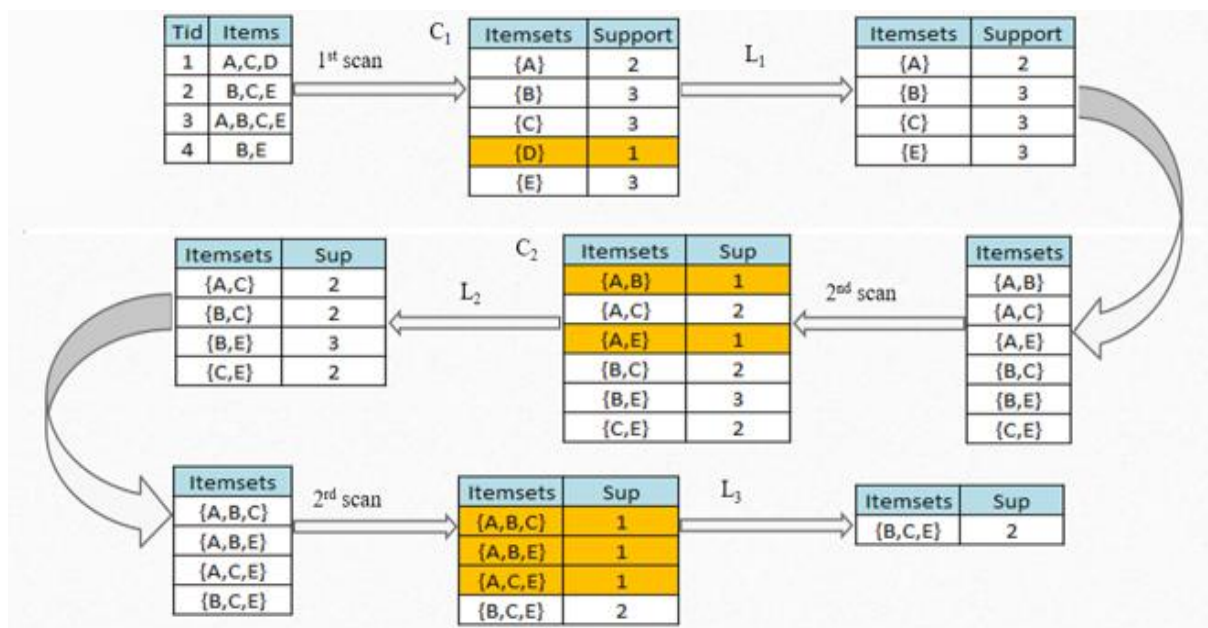
Hàm subset: tìm tất cả các ứng cử viên được chứa trong giao tác

Ví dụ: Giả sử tập các mục  $I = \{A, B, C, D, E\}$  và cơ sở dữ liệu giao dịch

*Bảng 1.1: Thang đo cảm nhận về tính hữu ích*

TID (giao dịch)	Các khoản mục (item)
1	{A, C, D}
2	{B, C, E}
3	{A, B, C, E}
4	{B, E}

Với ***minsup*** = 0.5 (tức tương đương 2 giao dịch). Khi thực hiện thuật toán Apriori trên ta có bảng trên. Các tập  $C_k$  và  $L_k$  sau khi thực hiện thuật toán Apriori trên tập D được thể hiện như sau:



**Hình 1.3 Ví dụ thuật toán Apriori**

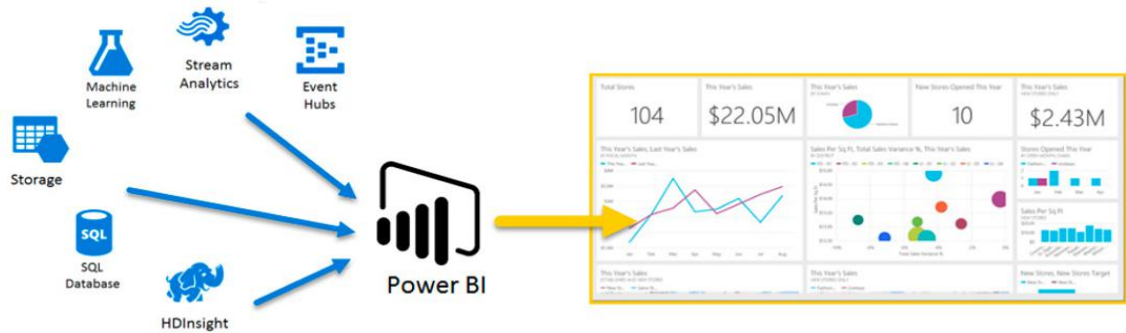


## CHƯƠNG 2

### GIỚI THIỆU CÁC CÔNG CỤ

#### 2.1. Tổng quan về công cụ trực quan hóa Power BI

##### *Khái niệm về Power BI*



**Hình 2.1** *Trực quan hóa dữ liệu với Power BI.*

Power BI là tập hợp các phần mềm, dịch vụ, ứng dụng và kết nối cho phép tích hợp dữ liệu từ nhiều nguồn cho mục đích tổng hợp, lập báo cáo, biểu diễn dữ liệu trực quan và xuất bản báo cáo dữ liệu lên môi trường web rất dễ dàng và thuận lợi.

Power BI cung cấp công cụ để thể hiện các dạng báo cáo (report) theo cách đơn giản hơn, trực quan hơn. Đồng thời cho phép chia sẻ dữ liệu một cách dễ dàng.

Power BI gồm 3 môi trường chính:

Power BI Desktop: Windows desktop application.

Power BI service: online SaaS (Software as a Service).

Power BI apps: Mobile devices (Android, Windows phones, iOS).



**Hình 2.2** *Các môi trường của Power BI.*

### ***Chức năng của BI***

- Đưa các thông tin của toàn bộ một doanh nghiệp lên một bảng biểu (Dashboard): Với Power BI, tất cả các dữ liệu quan trọng từ doanh nghiệp sẽ được theo dõi thông qua các ứng dụng cho phép hiển thị như là một bảng điều khiển hoặc những báo cáo tổng quát, biểu đồ theo thời gian.

- Tạo ra các báo cáo có tính tương tác cao: Power BI cung cấp cho bạn các công cụ để chuyển đổi, phân tích và hiển thị dữ liệu. hia sẽ báo cáo chỉ trong vài giây.

- Phân tích mức độ ổn định: Với Power BI, bạn rất dễ dàng để tạo ra những mô hình tái sử dụng dữ liệu (reuseable data) để đảm bảo sự nhất quán giữa các báo cáo, tiến hành các dự đoán, đưa ra các xu thế trong tương lai dựa vào các dữ liệu hiện tại và trong quá khứ. Những báo cáo vào các ứng dụng của bạn.

- Tích hợp ứng dụng của bạn với Power BI, các báo cáo tương tác và biểu đồ thời gian thực sẽ dễ dàng được chia sẻ đến những người dùng khác hay khách hàng của bạn.

### **2.2. Tổng quan về Microsoft SQL Server 2014**

- Microsoft SQL Server (MS SQL Server) là một hệ quản trị cơ sở dữ liệu (CSDL) quan hệ do hãng Microsoft phát triển sử dụng ngôn ngữ truy vấn Transact-SQL (T-SQL) để trao đổi dữ liệu giữa máy khách (Client) và máy chủ (Server) – Mô hình Client-Server. [4]

- Hiện nay MS SQL Server có thể thao tác với các CSDL có kích thước cực lớn, phục vụ nhiều máy khách cùng lúc và kết hợp ăn ý với các chương trình khác (như Microsoft Internet Information Server – IIS, Microsoft Team Foundation Server – TFS, Microsoft Visual Studio - VS,...). MS SQL Server chỉ chạy trên hệ điều hành windows của Microsoft. Nhưng việc kết nối đến MS SQL Server có thể thực hiện từ các hệ điều hành khác nhau dựa vào thư viện các nhà phát triển cung cấp.

- Các thành phần của một server: Database, Security, Server objects, Replication, Management.

- Database: thư mục chứa các CSDL của server này.

- Security: chứa các tài khoản người dùng (Logins), nhóm quyền (Server Roles) của server này.

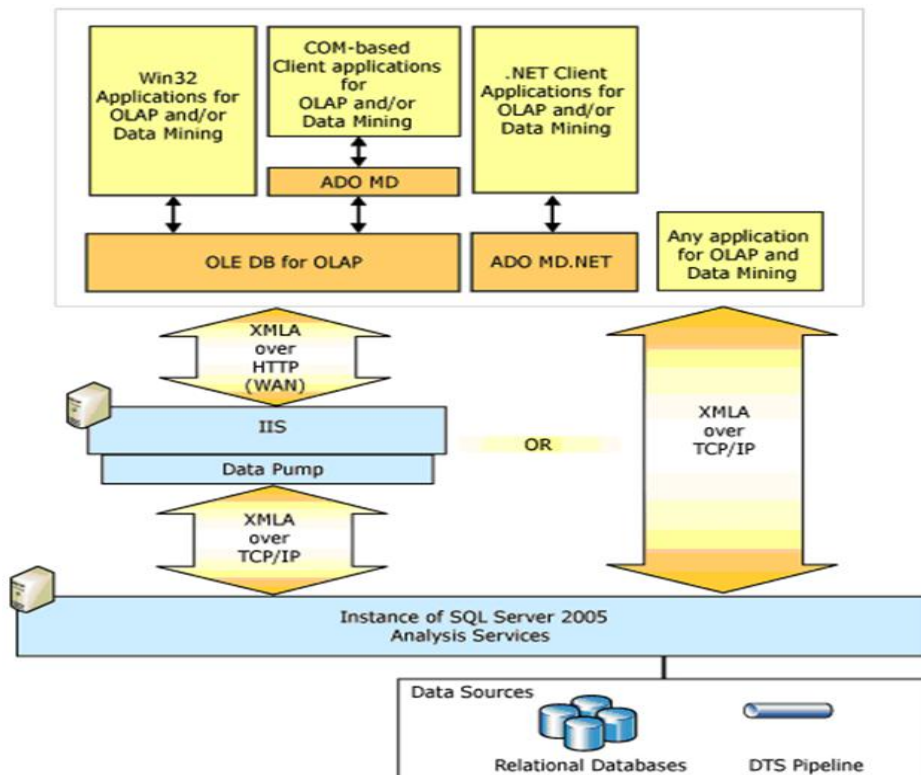
- Server objects Gồm Backup Devices, Linked Server, Database Server Tier, Server Tier, Client Tier, Trigger.

- Backup device: thực hiện sao lưu server này vào 1 file .bak
- Linked Server: kết nối đến nhiều CSDL khác. Các CSDL này có thể nằm trên một hệ quản trị CSDL khác MS SQL Server.
- Database Server Tier: Tầng CSDL. Tầng này chứa CSDL trên các hệ quản trị CSDL khác nhau: SQL Server, Oracle, Access, ODBC.
- Server Tier: Tại tầng này, SQL Server kết nối đến CSDL ở tầng CSDL thông qua các hàm mà OLE DB cung cấp. OLE DB là một tập các hàm (API) được thiết kế bởi Microsoft để truy cập dữ liệu từ nhiều nguồn khác nhau.
- Client Tier: Tầng các máy khách. Các máy trong tầng này kết nối đến MS SQL Server kết nối đến và thao tác với CSDL.
- Trigger.
- Replication Đây là một giải pháp tạo ra các bản sao dữ liệu giống hệt nhau, di chuyển các bản sao này đến những máy chủ khác nhau tại các vùng khác nhau mà vẫn đảm bảo đồng bộ hóa dữ liệu một cách tự động (các bản sao có cùng giá trị dữ liệu)  
Các vai trò trong Replication:
  - Publisher (người phát hành): là người lưu trữ CSDL nguồn, làm cho CSDL có khả năng nhân bản. + Distributor (người phân phối): Nhận và lưu những thay đổi rồi chuyển nó tới những người đăng ký (Subscriber).
  - Subscriber (người đăng ký): nhận những thay đổi dữ liệu, lưu trữ dữ liệu đã được nhân bản.
  - Management: Gồm: Policy Management, Extended Events, SQL Server Logs. Là các phương thức quản lý server, lưu trữ lịch sử sử dụng server.
- Bên cạnh thành phần cơ bản nhất của SQL Server là Database Engine có khả năng lưu trữ và xử lý dữ liệu ở các quy mô khác nhau như dạng bảng hay các kiểu kết nối thông dụng của Microsoft(ADO-Active Data Object, OLE DB, OLEDB – Open Database Connectivity) thì từ phiên bản SQL server 2005 trở đi thì bộ SQL Server đã được tích hợp nhiều gói dịch vụ hỗ trợ việc tích hợp và khai thác khả năng tiềm tàng của một cơ sở dữ liệu như SQL Integrating Services, SQL Server Analysis Services, Reporting Services.

### 2.3. Microsoft SQL Server Analysis Services (SSAS)

Analysis Services là một dịch vụ hỗ trợ mạnh mẽ việc phân tích, khai thác thông tin tiềm tàng bên trong một hệ quản trị cơ sở dữ liệu dựa vào khái niệm khối đa chiều (multiple-dimension cube) và khai phá dữ liệu.

Analysis Services là công cụ dễ sử dụng, tích hợp và linh động giúp định nghĩa các chiều, xây dựng các khối dữ liệu đa chiều và cung cấp các chương trình ứng dụng truy xuất tới các khối này.



**Hình 2.3 Mô hình kiến trúc SSAS.**

Microsoft SQL Server Analysis Services (SSAS) đáp ứng nhu cầu xử lý phân tích dữ liệu OLAP và các ứng dụng khai phá dữ liệu. Analysis Services hỗ trợ OLAP bằng cách cho phép bạn thiết kế, tạo ra, và quản lý các cấu trúc đa chiều có chứa dữ liệu được tổng hợp từ các nguồn dữ liệu khác, chẳng hạn như cơ sở dữ liệu quan hệ. Đối với các ứng dụng khai thác dữ liệu Analysis Services cho phép bạn thiết kế, tạo ra, và hình dung mô hình khai thác dữ liệu được xây dựng từ các nguồn dữ liệu khác bằng cách sử dụng một loạt các tiêu chuẩn các thuật toán khai thác dữ liệu. [5]

SSAS cho phép các ứng dụng hoạt động theo chế độ Client/Server. Thành phần server được triển khai như một dịch vụ của Microsoft Windows bao gồm một số thành phần thực hiện các chức năng:

- Phân tích câu lệnh nhận được từ client.
- Quản lý metadata.
- Xử lý tính toán.
- Lưu trữ dữ liệu.
- Tạo các khối tích hợp.
- Lập lịch truy vấn.
- Quản lý tài nguyên.

SSAS hỗ trợ nhiều dịch vụ trên cùng một máy chủ, mỗi dịch vụ là một đối tượng riêng biệt. Thành phần client truyền thông với server bằng giao thức XML for Analysis (XML/A) thông qua cơ chế truy vấn và trả lời tương tự như web service. Câu lệnh truy vấn sử dụng ngôn ngữ SQL, MDX (Multidimensional Expressions), DMX (Data Mining Extensions). Vài provider cũng được SSAS cung cấp để hỗ trợ các ngôn ngữ lập trình khác nhau.

#### **2.4. Môi trường triển khai ứng dụng (BIDS)**

Business Intelligence Development Studio (BIDS) là môi trường phát triển ứng dụng SSAS. Môi trường này được tích hợp với Microsoft Analysis Service. Phần lớn thời gian làm việc với SQL Server Data Mining là làm việc với BIDS. Nó là công cụ cho phép tổ chức và khai thác kho dữ liệu (xử lý phân tích trực tuyến) cũng như xây dựng mô hình khai phá dữ liệu rất dễ sử dụng và hiệu quả của Microsoft.

BIDS có hai chế độ hoạt động: Offline và Immediate.

- Làm việc với chế độ Immediate, người dùng kết nối trực tiếp và liên tục đến Analysis Service server. Khi tạo mới hoặc thay đổi các đối tượng đã có, các hoạt động này đều được thay đổi trên server. Mặc dù chế độ làm việc này rất trực quan và dễ hiểu. Tuy nhiên, nó cũng có nhiều hạn chế như các thao tác sai không thể phục hồi, các tác động không mong muốn từ các phiên làm việc khác, tính nhất quán giữa các ứng dụng sử dụng chung dữ liệu. Ngoài ra chế độ này chỉ bó hẹp trong một sơ sở dữ liệu nhất định.

- Trong chế độ Offline, các thay đổi trong dự án chỉ được lưu trữ ở client. Nó chỉ thành hiện thực khi quyết định triển khai ứng dụng lên server. Chế độ làm việc này phù hợp với việc thiết kế và chạy thử các mô hình trên máy chủ kiểm tra. Khi các mô hình đã được kiểm tra xác nhận thì mới được triển khai trên máy chủ.

## **CHƯƠNG 3**

### **PHÂN TÍCH PHẢN HỒI VỀ VIỆC SỬ DỤNG DỊCH VỤ ĐIỆN TẠI CÔNG TY TNHH MTV ĐIỆN LỰC ĐÀ NẴNG**

#### **3.1. Thực trạng và yêu cầu**

Hiện tại, công ty Điện lực Đà Nẵng thực hiện ghi nhận phản hồi cũng như mức độ hài lòng của khách hàng về các dịch vụ điện mà công ty cung cấp. Tính đến ngày 30/06/2019, dữ liệu bao gồm 7 cột: kênh liên lạc, mã khách hàng, loại khách hàng, phản hồi khách hàng, chủ đề yêu cầu, mức độ hài lòng khách hàng, quận và 44824 dòng.

Công ty muốn khai thác dữ liệu trên để đánh giá phân tích mức độ hài lòng của khách hàng đối với dịch vụ điện mà công ty cung cấp. Vì vậy, bài báo cáo này sẽ sử dụng dữ liệu mà công ty đã thu thập cùng với các công cụ: Power BI, Microsoft SQL Server 2014 và ứng dụng luật kết hợp để xây dựng mô hình khai phá dữ liệu nhằm phân loại các chủ đề phản hồi. Qua đó, phân tích mức độ hài lòng của khách hàng để quản lý cũng như hình thành chính sách chăm sóc hợp lý nhằm thỏa mãn nhu cầu khách hàng.

Dựa vào cơ sở dữ liệu về mức độ hài lòng của khách hàng khi sử dụng dịch vụ điện tại 5 chi nhánh Điện lực: Thanh Khê, Hòa Vang, Cẩm Lệ, Sơn Trà, Hải Châu thuộc Công ty TNHH MTV Điện Lực Đà Nẵng trong năm 2019. Mục đích của việc xây dựng mô hình khai phá dữ liệu nhằm để đánh giá mức độ hài lòng của khách hàng thông qua việc sử dụng dịch vụ chăm sóc khách hàng tại 5 chi nhánh Điện Lực ở Đà Nẵng. Từ đó hiểu được nhu cầu của khách hàng để đưa ra những giải pháp, chính sách chăm sóc khách hàng tốt nhất.

#### **3.2. Đề xuất quy trình**





### Hình 3.1 Đề xuất quy trình khai phá dữ liệu

#### 3.2.1 Thu thập dữ liệu

Dữ liệu được thu thập thông qua thống kê tại công ty TNHH MTV Điện Lực Đà Nẵng, dữ liệu gồm 44824 dòng.

	A	B	C	D	E	F	G	H
1	KENH_LIENLAC	MA_KH	LOAI_KH	PHANHAI_KH	CHUDE_YEUCAU	MUCDOHAILONG_KH	QUAN	
2	Điện thoại	PP09000130004	QL_TD	KH hỏi lịch cắt điện hôm nay. ĐTV đã trả lời	thay đổi thông tin đã đăng ký	Rất hài lòng	THANHKHE	
3	Điện thoại	PP09000806896	QL_TD	KH tra cứu tiền điện. Đã báo KH gọi lại sau vì không kiểm tra thông tin được.	thay đổi thông tin đã đăng ký	Không hài lòng	THANHKHE	
4	Điện thoại	PP09000114949	QL_TD	KH hỏi thời hạn thanh toán. ĐTV đã trả lời KH và hoàn tất phiếu	cấp điện trở lại khi KH đã tạm ngừng sử d	Rất hài lòng	THANHKHE	
5	Điện thoại	PP09000840626	QL_TD	KH không có yêu cầu và gác máy. ĐTV hoàn tất phiếu này	cấp điện trở lại khi KH đã tạm ngừng sử d	Rất không hài lòng	THANHKHE	
6	Website	PP09000902440	QL_TD	KH thông báo khi tải hóa đơn trên trang web thì chỉ có hóa đơn của tháng 2	thay đổi thông tin đã đăng ký	Bình thường	THANHKHE	
7	Điện thoại	PP09000815141	QL_TD	KH nhờ cung cấp MKH, ĐTV đã cung cấp thông tin	thay đổi thông tin đã đăng ký	Hài lòng	THANHKHE	
8	Email	PP09000822778	CN_XD	KH hỏi cách tra cứu thông tin trên trang web cskh.vpc.vn. ĐTV đã hướng dẫn	thay đổi thông tin đã đăng ký	Không hài lòng	THANHKHE	
9	Điện thoại	PP09000915358	QL_TD	KH tra cứu thông tin, ĐTV đã trả lời	thay đổi thông tin đã đăng ký	Rất hài lòng	THANHKHE	
10	Điện thoại	PP09000804834	QL_TD	KH trong khu vực bảo cách đây 3 ngày ĐL có đến phát quang hành lang tuyến	cấp điện trở lại khi KH đã tạm ngừng sử d	Không hài lòng	THANHKHE	
11	Điện thoại	PP09000832819	QL_TD	KH muốn đăng ký SĐT nhận thông báo tiền điện hàng tháng qua SMS. SĐT đã	thay đổi thông tin đã đăng ký	Bình thường	THANHKHE	
12	Điện thoại	PP09000907282	TN_KS_NH	KH tra cứu tên đăng nhập và mật khẩu truy cập web CSKH. ĐTV đã trả lời KH	cấp điện trở lại khi KH đã tạm ngừng sử d	Hài lòng	THANHKHE	
13	Điện thoại	PP09000901569	QL_TD	KH hỏi lịch sự cố ở khu vực. ĐTV đã trả lời KH.	cấp điện trở lại khi KH đã tạm ngừng sử d	Hài lòng	THANHKHE	
14	Điện thoại	PP09000820887	QL_TD	KH đăng ký qua Website. KH có nhu cầu thay đổi SĐT nhận thông tin từ ĐL: 09	thay đổi thông tin đã đăng ký	Rất hài lòng	THANHKHE	
15	Điện thoại	PP09000894640	QL_TD	KH tra cứu thông tin, ĐTV đã trả lời	thay đổi thông tin đã đăng ký	Rất hài lòng	THANHKHE	
16	Điện thoại	PP09000130795	QL_TD	Khách hàng tra cứu lịch cắt điện. CSKH đã cung cấp thông tin lịch cắt điện ch	thay đổi thông tin đã đăng ký	Bình thường	THANHKHE	
17	Điện thoại	PP09000804002	QL_TD	KH tra cứu lịch mất điện sự cố. ĐTV đã trả lời khách hàng.	giải đáp kiến nghị của khách hàng	Không hài lòng	THANHKHE	
18	Điện thoại	PP09000908889	QL_TD	KH hỏi thông tin lịch cắt điện. ĐTV đã trả lời cho KH.	thay đổi thông tin đã đăng ký	Rất hài lòng	THANHKHE	
19	Điện thoại	PP09000912253	QL_TD	KH báo không nhận được thông báo tiền điện qua email: nguyenvquydr@gmail	thay đổi thông tin đã đăng ký	Rất không hài lòng	THANHKHE	
20	Điện thoại	PP09000840633	TN_KS_NH	KH hỏi thông tin đăng nhập vào website. ĐTV đã trả lời.	thay đổi thông tin đã đăng ký	Rất hài lòng	THANHKHE	
21	Điện thoại	PP09000893195	QL_TD	KH hỏi MK web cskh. ĐTV đã trả lời.	thay đổi thông tin đã đăng ký	Hài lòng	THANHKHE	
22	Điện thoại	PP09000881448	QL_TD	KH tra cứu thông tin phiếu.2534039. ĐTV trả lời ĐL đã tiếp nhận thông tin và	thay đổi thông tin đã đăng ký	Hài lòng	THANHKHE	
23	Điện thoại	PP09000896589	QL_TD	KH báo nhận được phiếu thông báo ngừng cung cấp điện trong vòng 24h nếu	thay đổi hình thức thanh toán tiền điện	Không hài lòng	THANHKHE	
24	Điện thoại	PP09000135250	QL_TD	KH hỏi lịch cắt điện. ĐTV đã trả lời KH.	cấp điện trở lại khi KH đã tạm ngừng sử d	Bình thường	THANHKHE	
25	Điện thoại	PP09000918786	CN_XD	KH có nhu cầu tạm ngưng cấp điện trong khoảng 3 ngày để xử lý công việc n	cấp điện trở lại khi KH đã tạm ngừng sử d	Rất hài lòng	THANHKHE	
26	Điện thoại	PP09000882218	QL_TD	KH hỏi cách đăng nhập trên App CSKH ĐTV đã hướng dẫn KH.	cấp điện trở lại khi KH đã tạm ngừng sử d	Bình thường	THANHKHE	
27	Điện thoại	PP09000830874	QL_TD	ĐTV đã giải đáp.	giải đáp kiến nghị của khách hàng	Không hài lòng	THANHKHE	

### Hình 3.2 Dữ liệu thu thập được chưa qua xử lý

#### 3.2.2 Tiền xử lý dữ liệu

- Loại bỏ các cột: Mã khách hàng, phản hồi khách hàng.
- Thêm cột ID khách hàng với số thứ tự từ ID001 đến ID44824.

	A	B	C	D	E	F
1	ID	KENH_LIENLAC	LOAI_KH	CHUDE_YEUCAU	MUCDOHAILONG_KH	QUAN
2	ID001	Điện thoại	QL_TD	thay đổi thông tin đã đăng ký	Rất hài lòng	THANHKHE
3	ID002	Điện thoại	QL_TD	thay đổi thông tin đã đăng ký	Không hài lòng	THANHKHE
4	ID003	Điện thoại	QL_TD	cấp điện trở lại khi KH đã tạm ngừng sử dụng điện	Rất hài lòng	THANHKHE
5	ID004	Điện thoại	QL_TD	cấp điện trở lại khi KH đã tạm ngừng sử dụng điện	Rất không hài lòng	THANHKHE
6	ID005	Website	QL_TD	thay đổi thông tin đã đăng ký	Bình thường	THANHKHE
7	ID006	Điện thoại	QL_TD	thay đổi thông tin đã đăng ký	Hài lòng	THANHKHE
8	ID007	Email	CN_XD	thay đổi thông tin đã đăng ký	Không hài lòng	THANHKHE
9	ID008	Điện thoại	QL_TD	thay đổi thông tin đã đăng ký	Rất hài lòng	THANHKHE
10	ID009	Điện thoại	QL_TD	cấp điện trở lại khi KH đã tạm ngừng sử dụng điện	Không hài lòng	THANHKHE
11	ID010	Điện thoại	QL_TD	thay đổi thông tin đã đăng ký	Bình thường	THANHKHE
12	ID011	Điện thoại	TN_KS_NH	cấp điện trở lại khi KH đã tạm ngừng sử dụng điện	Hài lòng	THANHKHE
13	ID012	Điện thoại	QL_TD	cấp điện trở lại khi KH đã tạm ngừng sử dụng điện	Hài lòng	THANHKHE
14	ID013	Điện thoại	QL_TD	thay đổi thông tin đã đăng ký	Rất hài lòng	THANHKHE
15	ID014	Điện thoại	QL_TD	thay đổi thông tin đã đăng ký	Rất hài lòng	THANHKHE
16	ID015	Điện thoại	QL_TD	thay đổi thông tin đã đăng ký	Bình thường	THANHKHE
17	ID016	Điện thoại	QL_TD	giải đáp kiến nghị của khách hàng	Không hài lòng	THANHKHE
18	ID017	Điện thoại	QL_TD	thay đổi thông tin đã đăng ký	Rất hài lòng	THANHKHE
19	ID018	Điện thoại	QL_TD	thay đổi thông tin đã đăng ký	Rất không hài lòng	THANHKHE
20	ID019	Điện thoại	TN_KS_NH	thay đổi thông tin đã đăng ký	Rất hài lòng	THANHKHE
21	ID020	Điện thoại	QL_TD	thay đổi thông tin đã đăng ký	Hài lòng	THANHKHE
22	ID021	Điện thoại	QL_TD	thay đổi thông tin đã đăng ký	Hài lòng	THANHKHE
23	ID022	Điện thoại	QL_TD	thay đổi hình thức thanh toán tiền điện	Không hài lòng	THANHKHE

### Hình 3.3. Tiền xử lý dữ liệu

### 3.2.3 Thiết kế dữ liệu

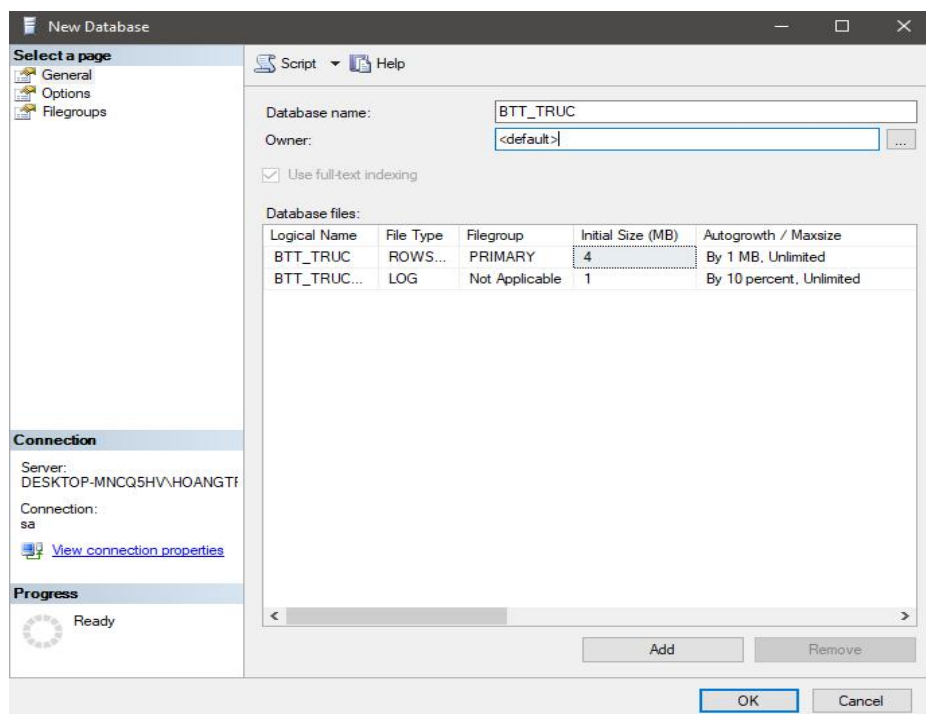
Bảng 3.1: Mô tả dữ liệu

Tên trường	Kiểu dữ liệu	Kích thước	Mô tả
ID	varchar	7	ID
LOẠI_KH	varchar	100	Loại khách hàng
KENH_LIENLAC	varchar	50	Kênh liên lạc
CHUDE_YEUCAU	Nvarchar	200	Chủ đề yêu cầu
MUCDOHAILONG_KH	Nvarchar	100	Mức độ hài lòng khách hàng
QUAN	varchar	30	Quận

### 3.2.4 Cài đặt dữ liệu vào hệ thống

- Xây dựng một cơ sở dữ liệu trên hệ quản trị Microsoft SQL Server 2014, thực hiện thu thập dữ liệu rồi nạp vào cơ sở dữ liệu thông qua chức năng Import của Microsoft SQL Server 2014.

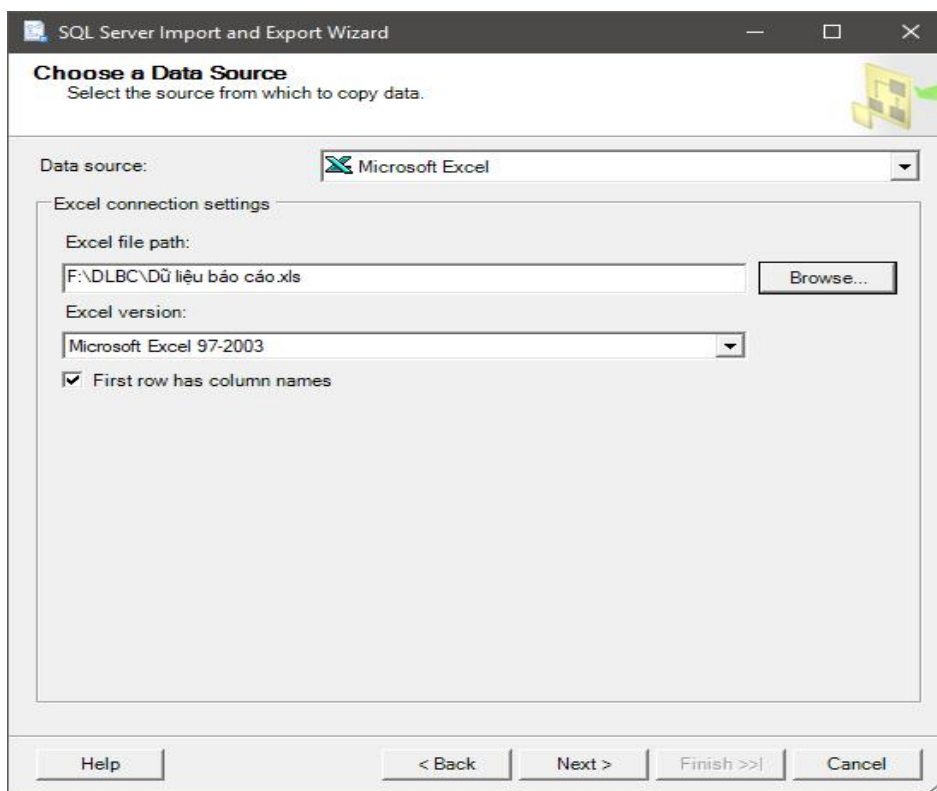
- Tạo Database mới: Vào Database ➔ New Database ➔ Database name: BTT\_TRUC.



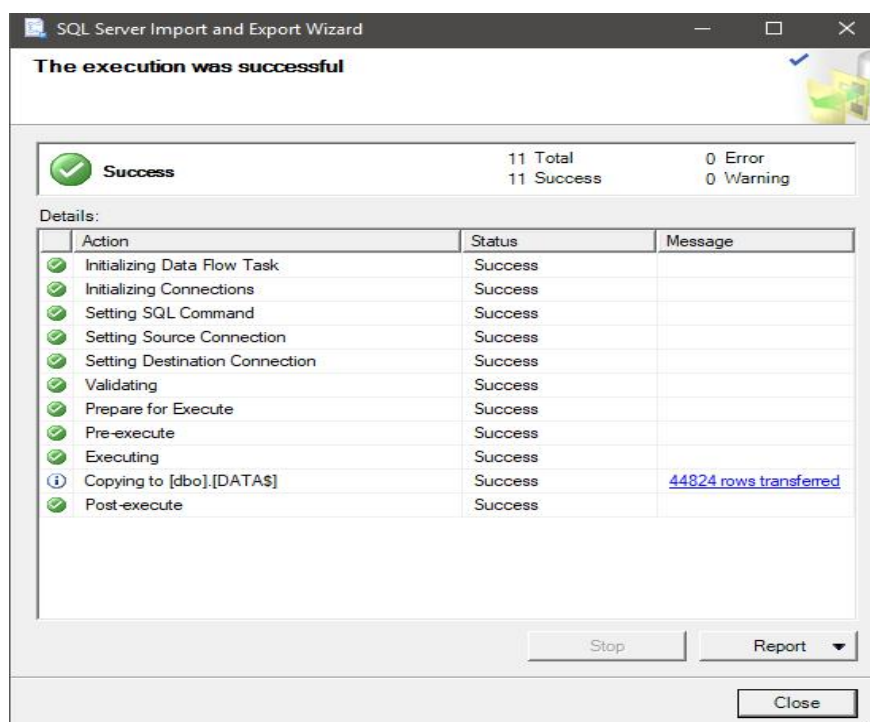


### ***Hình 3.4 Tạo database mới trong Microsoft SQL Server 2014.***

- Import dữ liệu vào Microsoft SQL Server 2014: Kích phải chuột vào database BTT\_TRUC → Task → Import Data → Next



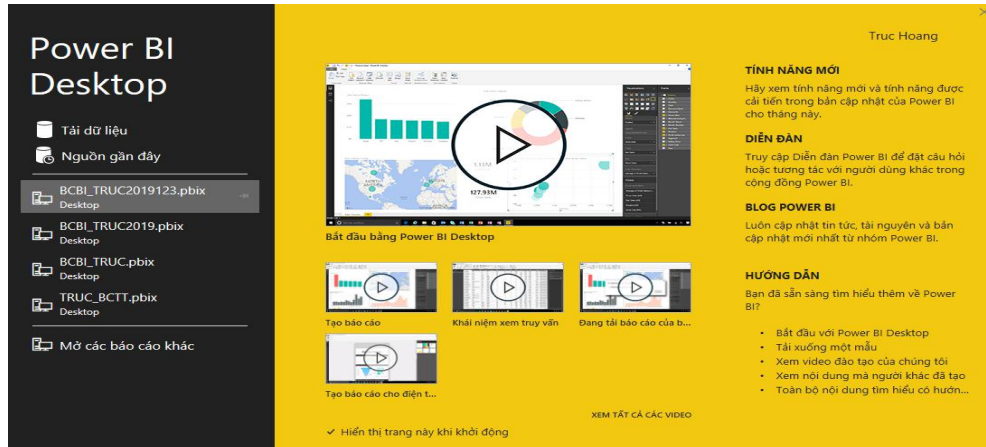
### ***Hình 3.5 Import dữ liệu vào database.***



**Hình 3.6 Import dữ liệu thành công.**

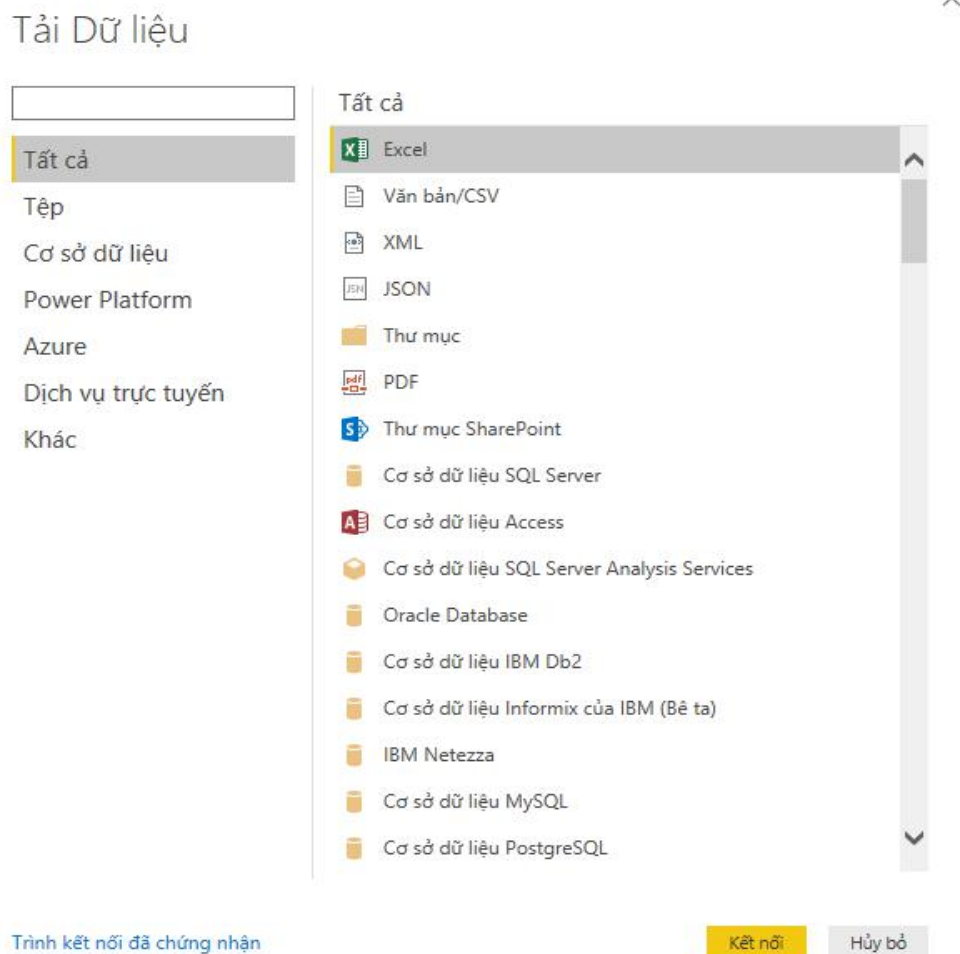
### 3.2.5 Trực quan hóa dữ liệu

- Sử dụng công cụ Power BI để trực quan hóa dữ liệu.



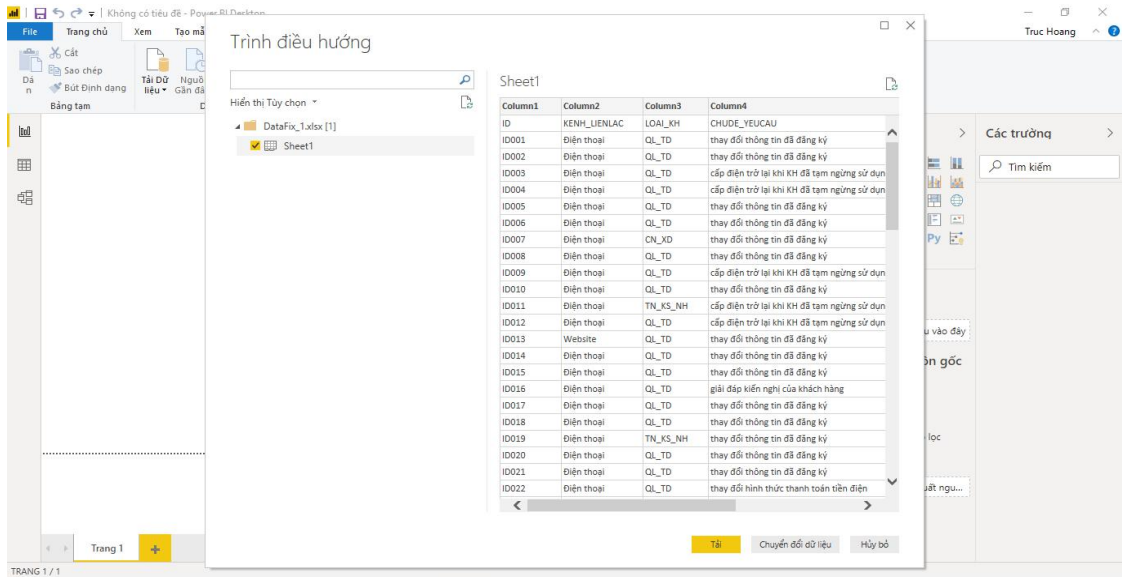
**Hình 3.7 Sử dụng công cụ Power BI.**

- Để thêm dữ liệu vào Power BI: Tải dữ liệu → Excel → Kết nối.



### Hình 3.8 Tải dữ liệu vào Power BI.

- Sau khi tải dữ liệu, giao diện Power BI sẽ hiện lên trình điều hướng.

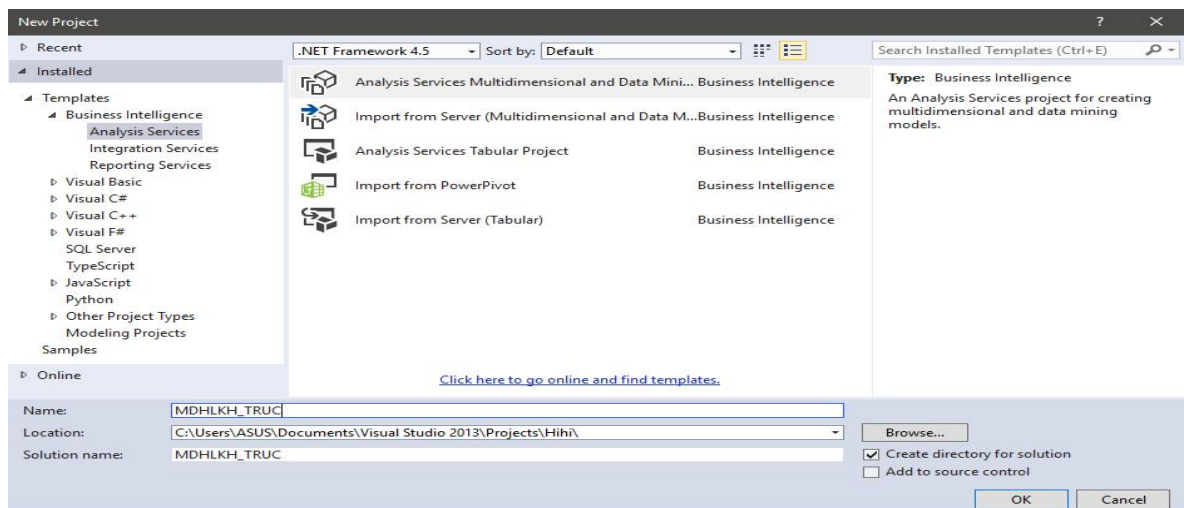


Hình 3.9 Trình điều hướng sau khi tải dữ liệu trong Power BI.

### 3.2.6 Xây dựng mô hình khai phá dữ liệu trên BIDS

#### Bước 1: Tạo Data Source

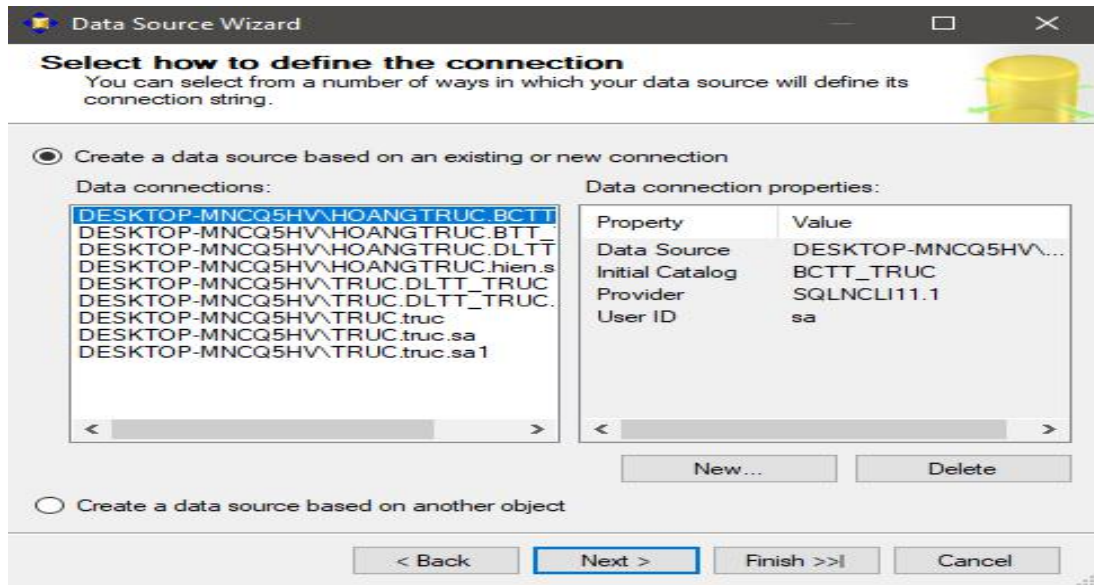
- Tạo project mới: File → New → Project → Xuất hiện hộp thoại New Project.
- Đặt tên project là: MDHLKH



Hình 3.10 Quy trình xây dựng mô hình luật kết hợp.

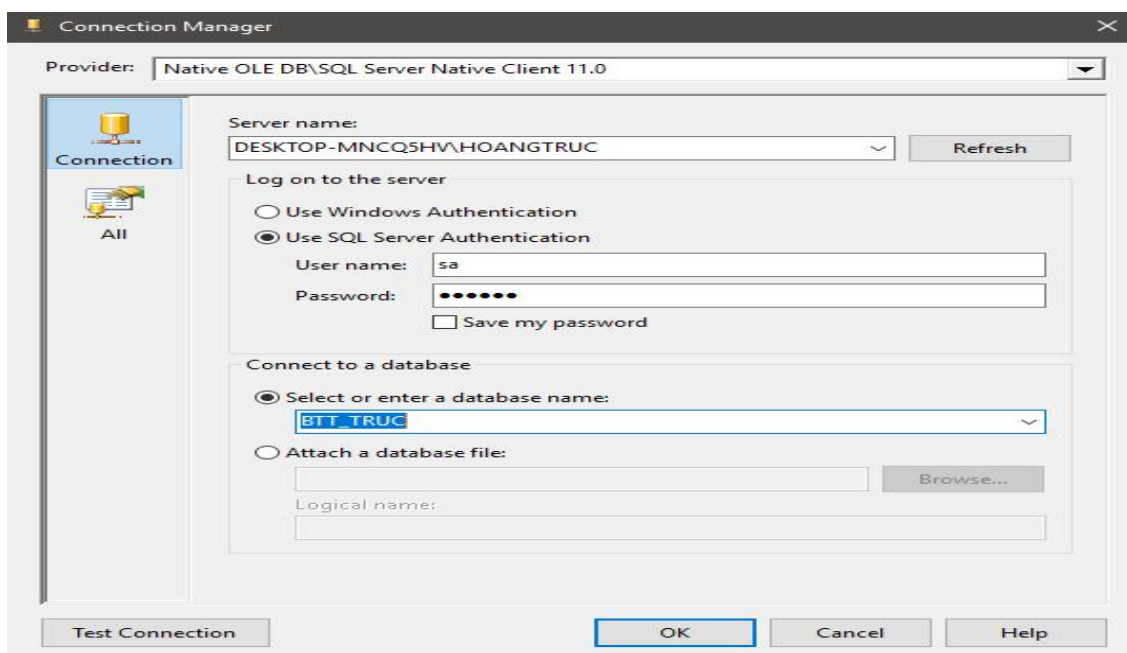
- Tại màn hình Welcome to Data Source Wizard, nhấn Next để chuyển sang cửa sổ Select how to define the connection. Tại đây, định nghĩa một datasource bằng cách

tạo một connection. Kiểm tra checkbox Create a data source based on an existing or new connection được chọn và nhấn New.



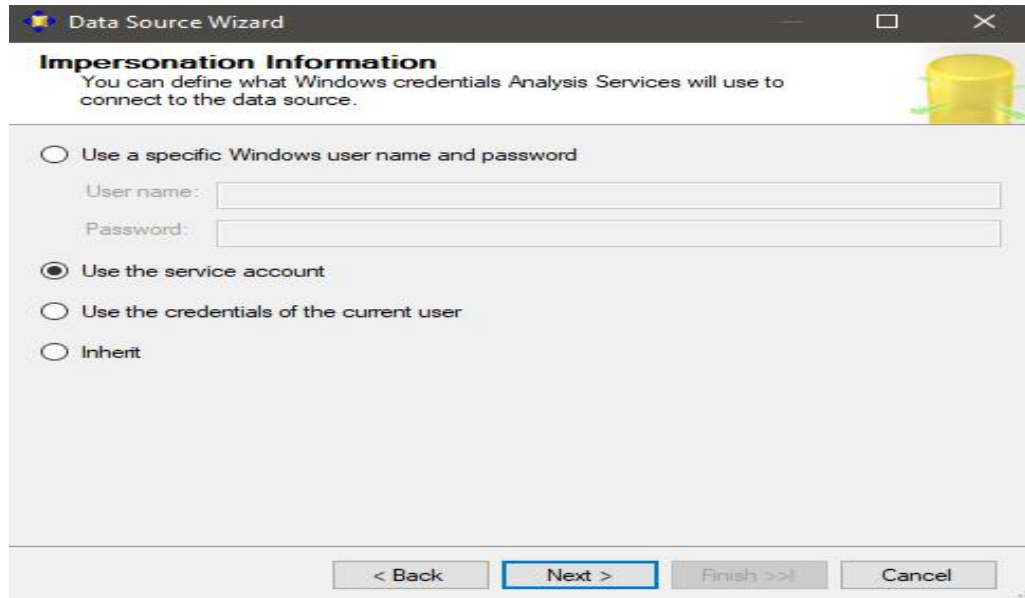
**Hình 3.11 Tạo Data Source Wizard.**

- Trong danh sách Provider, chọn Native OLE DB/SQL Server Native Client 11.0.
- Trong Server Name, gõ vào DESKTOP-MNCQ5HV\HOANGTRUC.
- Trong danh sách Select or enter a database name, chọn BTT\_TRUC.
- Nhấn Test Connection để kiểm tra kết nối đến SQL Server.
- Chọn OK và nhấn Next.



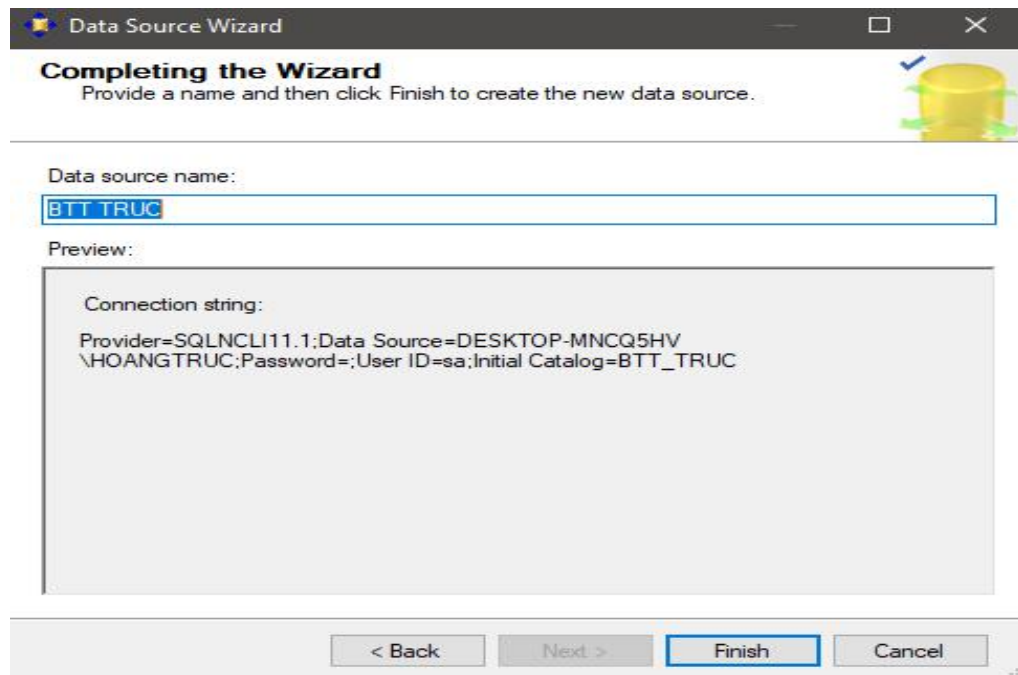
### Hình 3.12 Tạo Data Source

- Tại cửa sổ Impersonation Information, ta sẽ thiết lập Security Credential cho SSAS để kết nối tới Data Sources, chọn Use the service account và nhấn Next.



### Hình 3.13 Tạo Data Source

- Tại màn hình Completing the wizard, thiết lập tên của Data Source là BTT\_TRUC và chọn Finish.

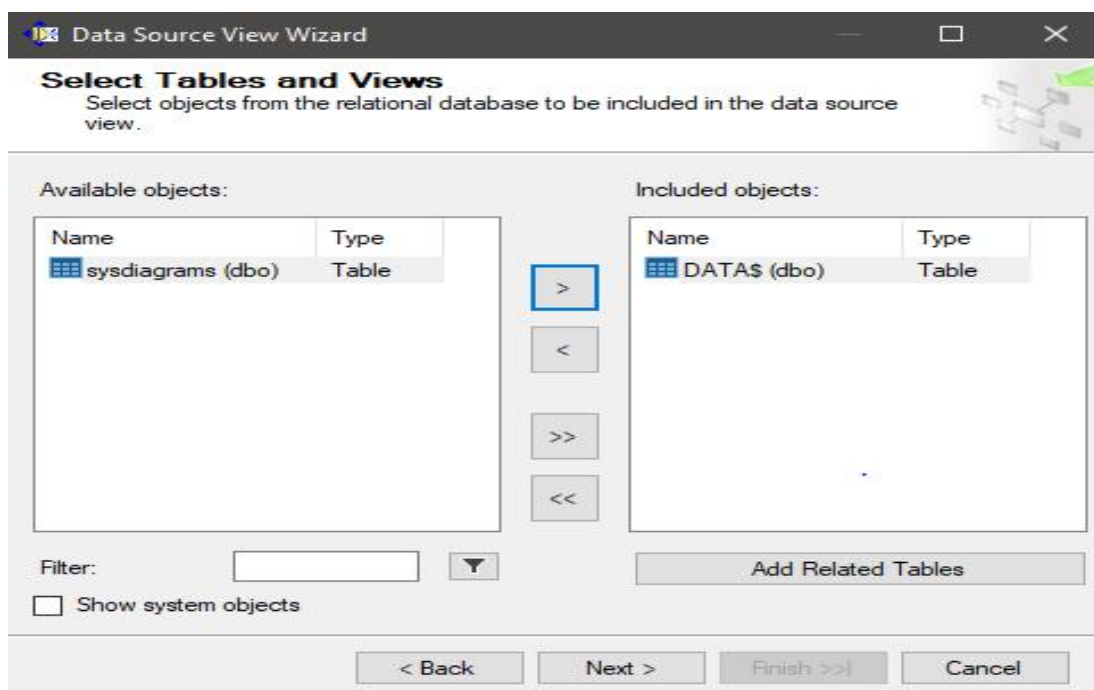


### Hình 3.14 Tạo Data Source

## Bước 2: Tạo Data Source View

- Trong khung Solution Explorer, kích phải chuột Data Source View, chọn New Data Source Views. Tại mục Relational Data Sources, chọn BTT\_TRUC đã tạo như trên, rồi chọn Next.

- Tại hộp thoại Select Tables and Views, chọn bảng DATA để thêm bảng này vào danh sách Included objects, rồi nhấn Next.



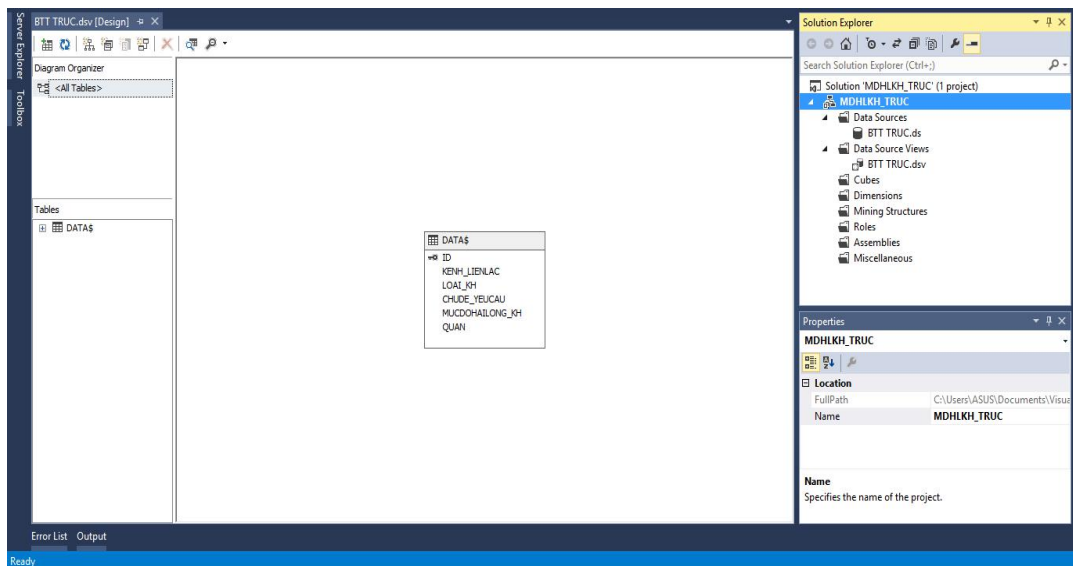
**Hình 3.15 Tạo Data Source View Wizard**

- Tại thuộc tính Name, ta thiết lập DATA và nhấn Finish để hoàn tất việc định nghĩa một Data Source View.

- Data Source View sẽ hiển thị tại Data Source View Designer trong BIDS và gồm các thành phần sau:

- Diagram Pane: thể hiện mối quan hệ giữa các bảng dữ liệu trực quan.
  - Tables Pane: các bảng dữ liệu và schema được thể hiện theo tree view.
  - Diagram Organizer Pane: hỗ trợ tạo các subdiagram mà thông qua đó ta có thể có được các subsets của data source view.
- Sau khi hoàn tất bước này, ta sẽ có khung nhìn toàn diện về bảng dữ liệu.

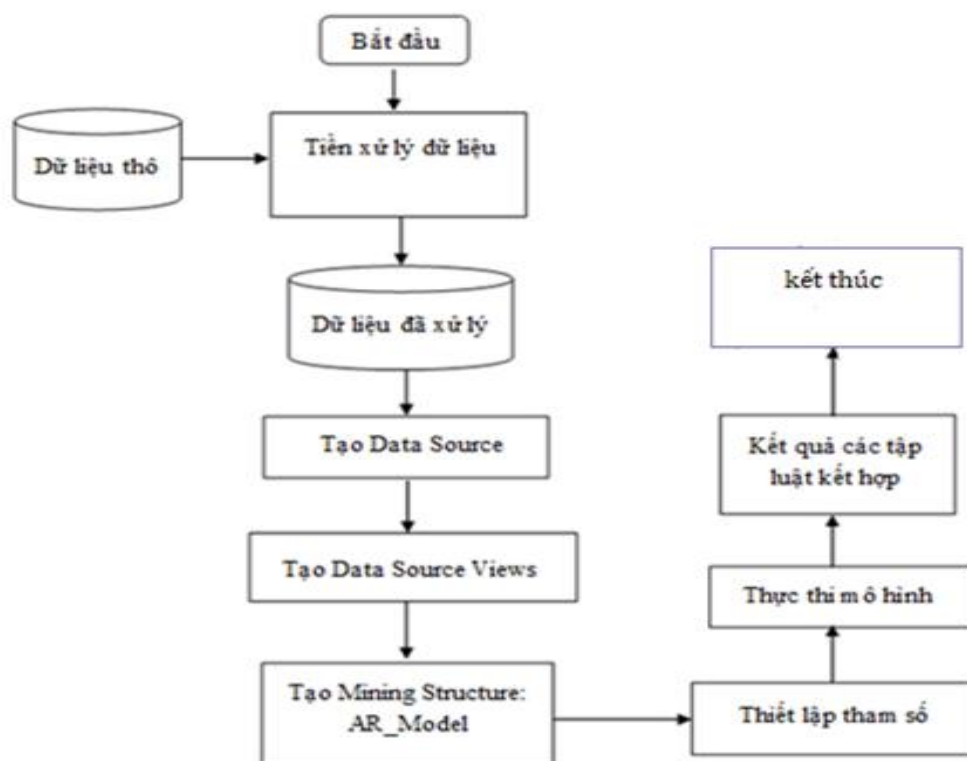




**Hình 3.16 Tạo Data Source View**

### **3.2.7 Ứng dụng luật kết hợp phân tích kết quả**

#### **Quy trình xây dựng mô hình luật kết hợp**



**Hình 3.17 Quy trình xây dựng mô hình luật kết hợp.**

### ***Dữ liệu thô***

- Một cột khóa duy nhất: Mỗi mô hình phải có một cột kiểu số hoặc kiểu chuỗi để xác định duy nhất một bản ghi.

- Một cột dự đoán duy nhất: Một mô hình kết hợp chỉ có thể có một cột dự đoán. Thông thường nó là cột khóa của bảng lồng nhau. Các giá trị phải rời rạc hoặc được rời rạc.

- Những cột đầu vào: Các cột đầu vào phải là rời rạc. Các dữ liệu đầu vào cho một mô hình kết hợp thường được chứa trong hai bảng. Chúng ta có thể nhập dữ liệu này vào mô hình bằng cách sử dụng một bảng lồng nhau.

### ***Dữ liệu đã xử lý***

- Dữ liệu sử dụng là 5 mức đánh giá độ hài lòng của khách hàng theo 16 chủ đề yêu cầu thuộc 5 chi nhánh Điện lực trên địa bàn Thành phố Đà Nẵng.

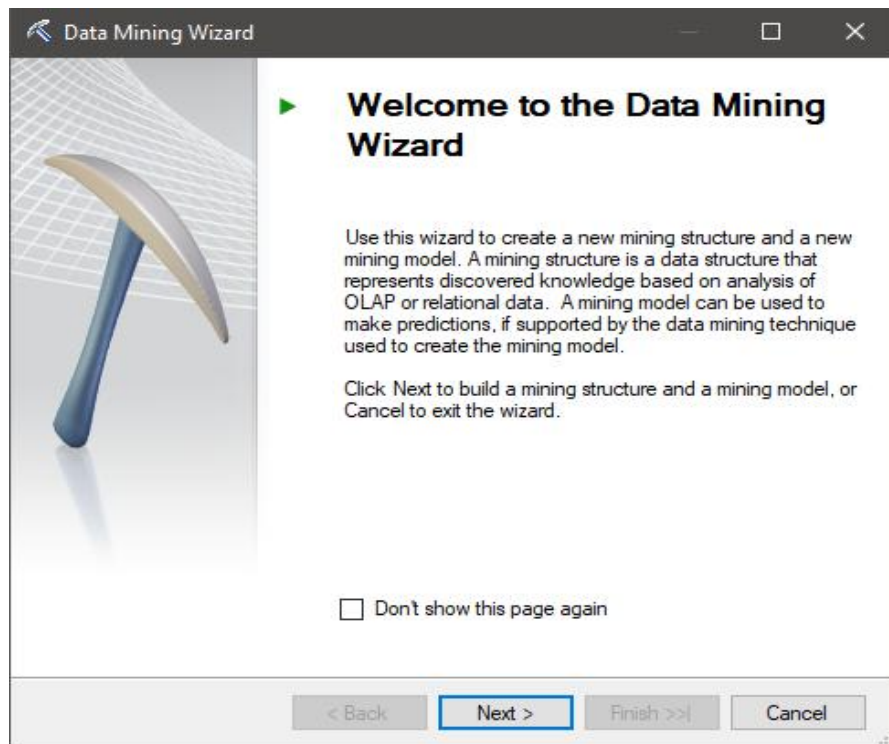
	A	B	C	D	E	F
1	ID	KENH_LIENLAC	LOAI_KH	CHUDE_YEUCAU	MUCDOHAILONG_KH	QUAN
2	ID001	Điện thoại	QL_TD	thay đổi thông tin đã đăng ký	Rất hài lòng	THANHKHE
3	ID002	Điện thoại	QL_TD	thay đổi thông tin đã đăng ký	Không hài lòng	THANHKHE
4	ID003	Điện thoại	QL_TD	cấp điện trở lại khi KH đã tạm ngừng sử dụng điện	Rất hài lòng	THANHKHE
5	ID004	Điện thoại	QL_TD	cấp điện trở lại khi KH đã tạm ngừng sử dụng điện	Hài lòng	THANHKHE
6	ID005	Điện thoại	QL_TD	thay đổi thông tin đã đăng ký	Không hài lòng	THANHKHE
7	ID006	Điện thoại	QL_TD	cấp điện trở lại khi KH đã tạm ngừng sử dụng điện	Không hài lòng	THANHKHE
8	ID007	Điện thoại	CN_XD	thay đổi hình thức thanh toán tiền điện	Rất không hài lòng	THANHKHE
9	ID008	Điện thoại	QL_TD	cấp điện trở lại khi KH đã tạm ngừng sử dụng điện	Không hài lòng	THANHKHE
10	ID009	Điện thoại	QL_TD	cấp điện trở lại khi KH đã tạm ngừng sử dụng điện	Hài lòng	THANHKHE
11	ID010	Điện thoại	QL_TD	cấp điện trở lại khi KH đã tạm ngừng sử dụng điện	Không hài lòng	THANHKHE
12	ID011	Điện thoại	TN_KS_NH	cấp điện trở lại khi KH đã tạm ngừng sử dụng điện	Hài lòng	THANHKHE
13	ID012	Điện thoại	QL_TD	cấp điện trở lại khi KH đã tạm ngừng sử dụng điện	Hài lòng	THANHKHE
14	ID013	Website	QL_TD	thay đổi thông tin đã đăng ký	Rất hài lòng	THANHKHE
15	ID014	Điện thoại	QL_TD	thay đổi hình thức thanh toán tiền điện	Rất không hài lòng	THANHKHE
16	ID015	Điện thoại	QL_TD	thay đổi thông tin đã đăng ký	Bình thường	THANHKHE
17	ID016	Điện thoại	QL_TD	giải đáp kiến nghị của khách hàng	Không hài lòng	THANHKHE
18	ID017	Điện thoại	QL_TD	thay đổi thông tin đã đăng ký	Rất hài lòng	THANHKHE

***Hình 3.18 Dữ liệu mức độ hài lòng của khách hàng.***

### ***Tạo data mining structures AR\_model***

- Kích phải chuột vào Mining Structures → chọn New Mining Structure.

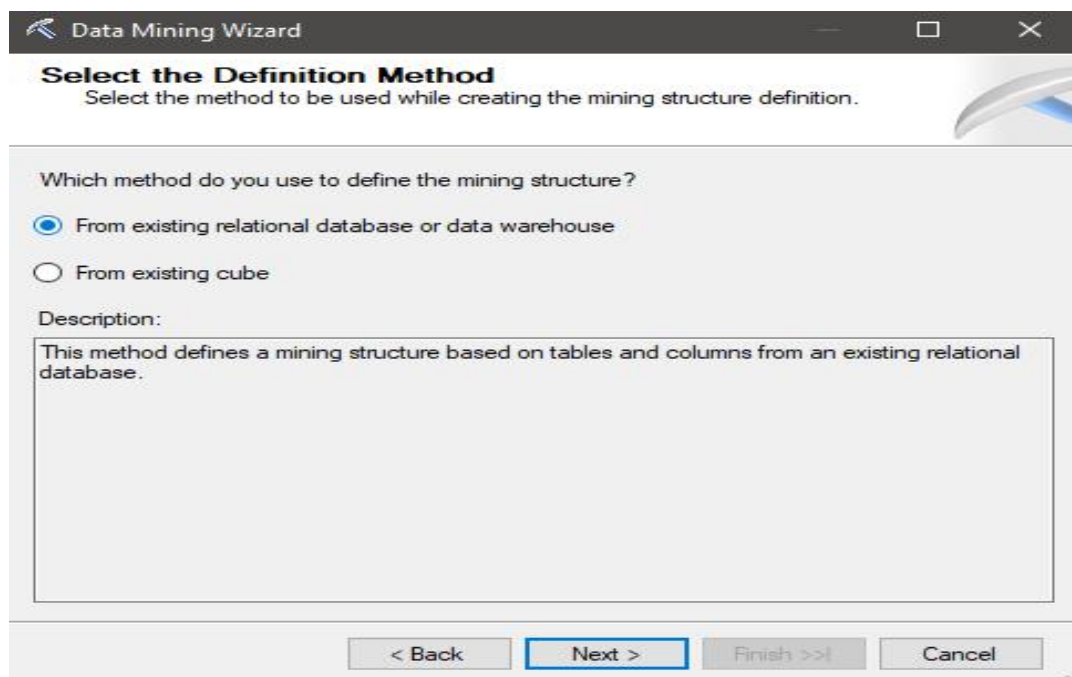




**Hình 3.19** *Giao diện tạo Mining Structures.*

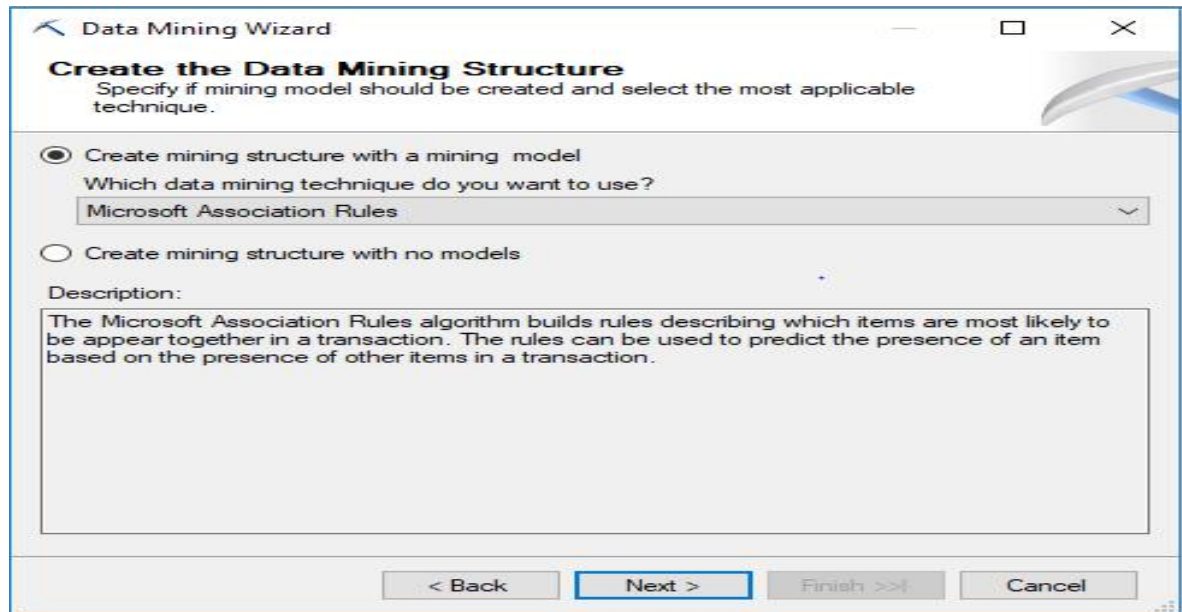
- Data Mining Wizard sẽ tạo ra cấu trúc và mô hình khai phá dữ liệu. Sau đó chọn phương pháp để xử lý mô hình dựa vào các bảng và các cột từ cơ sở dữ liệu quan hệ, chọn From existing relational database or data warehouse.

- Sau đó chọn Next.



**Hình 3.20** *Tạo Mining Structure*

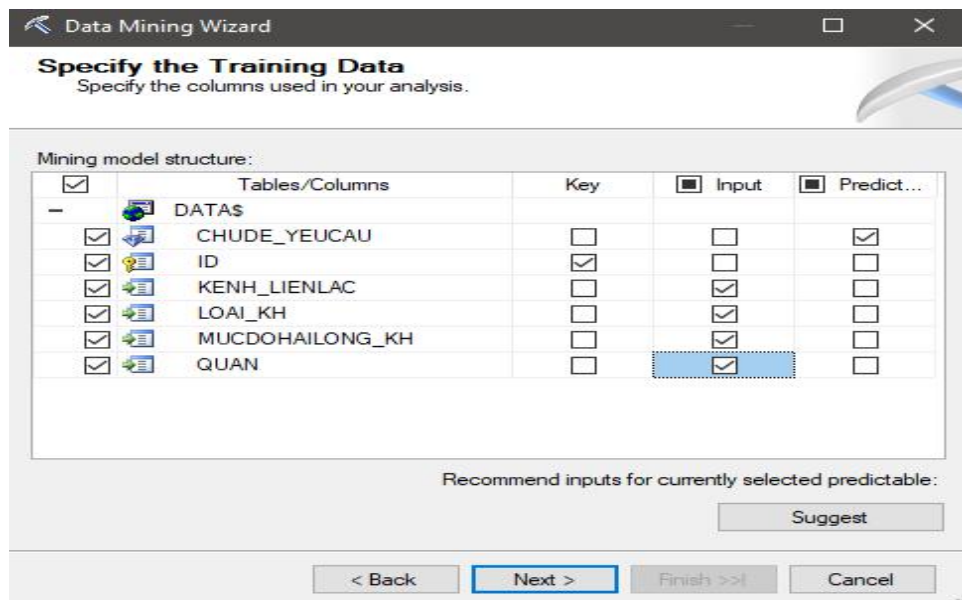
- Hộp thoại liệt kê tất cả những thuật toán khai phá dữ liệu, ở đây chọn thuật toán Microsoft Association Rules.



**Hình 3.21 Lựa chọn mô hình luật kết hợp**

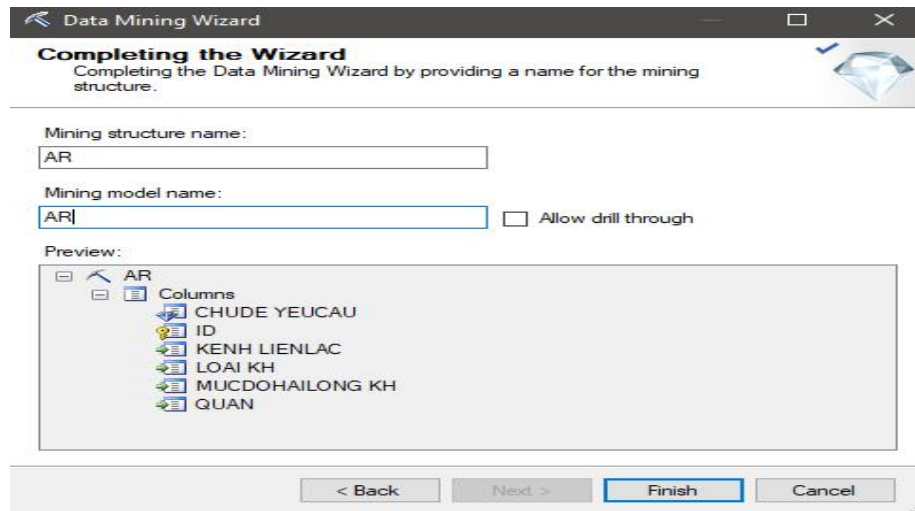
- Chọn thuộc tính để Training:

Chọn các thuộc tính Input: KENH\_LIENLAC, QUAN, LOAI\_KH, MUCDOHAILONG\_KH, thuộc tính CHUDE\_YEUCAU chọn Predictable.



**Hình 3.22 Chọn thuộc tính để Training.**

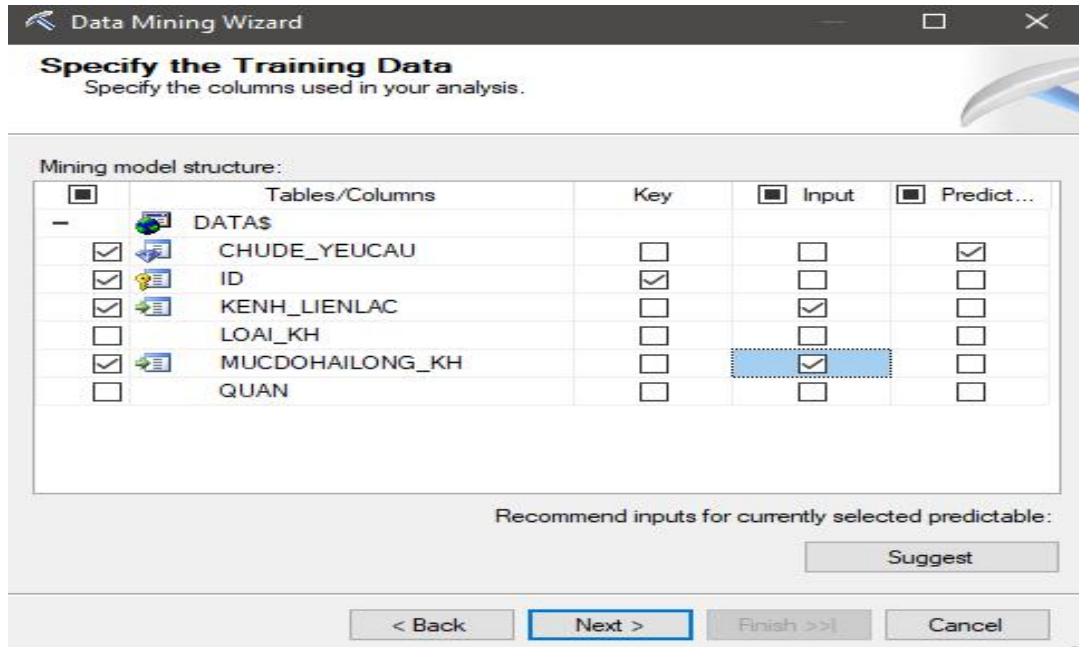
- Bước tiếp theo đặt lại tên cho mô hình khai phá dữ liệu là: AR
- Sau đó chọn Finish.



**Hình 3.23 Đặt tên mô hình luật kết hợp**

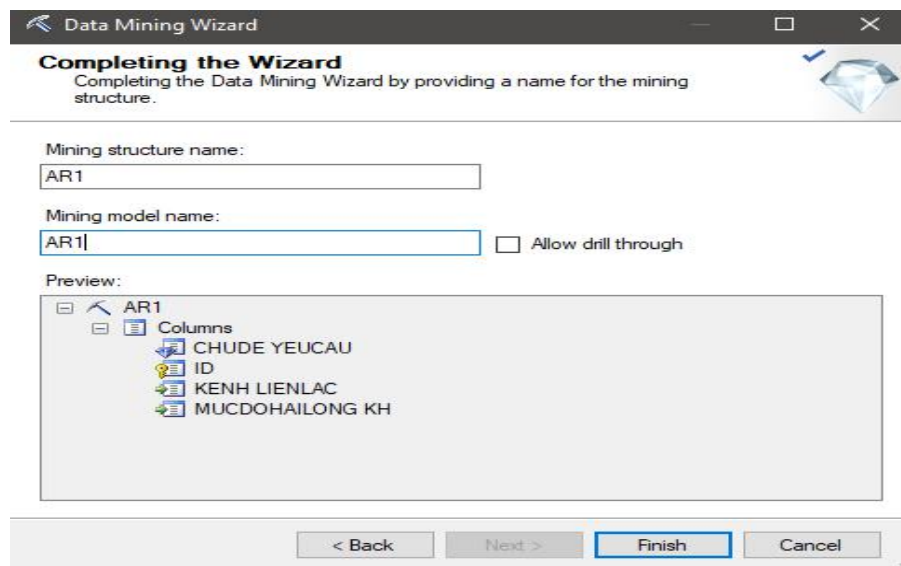
- Chọn các thuộc tính để Training:

Chọn các thuộc tính Input: KENH\_LIENLAC, MUCDOHAILONG\_KH, thuộc tính CHUDE\_YEUCAU chọn Predictable, ignore thuộc tính LOAI\_KH.



**Hình 3.24 Chọn thuộc tính để Training.**

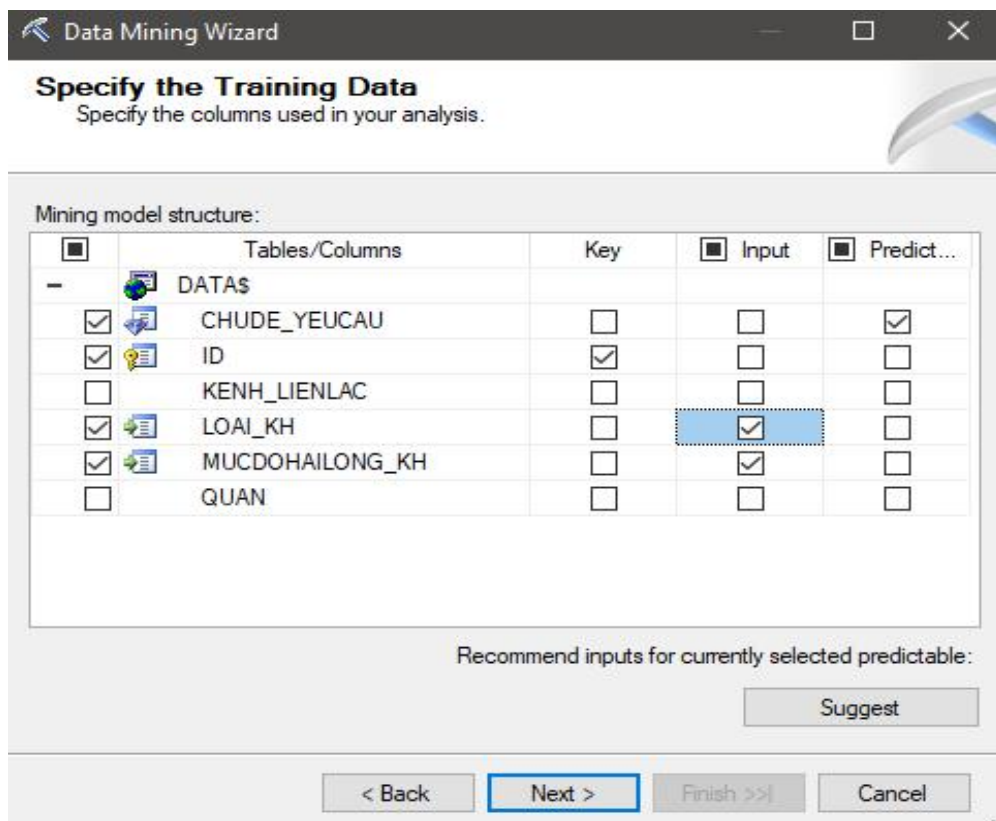
- Đặt tên cho mô hình này là AR1 và chọn Finish.



**Hình 3.25 Đặt tên mô hình luật kết hợp.**

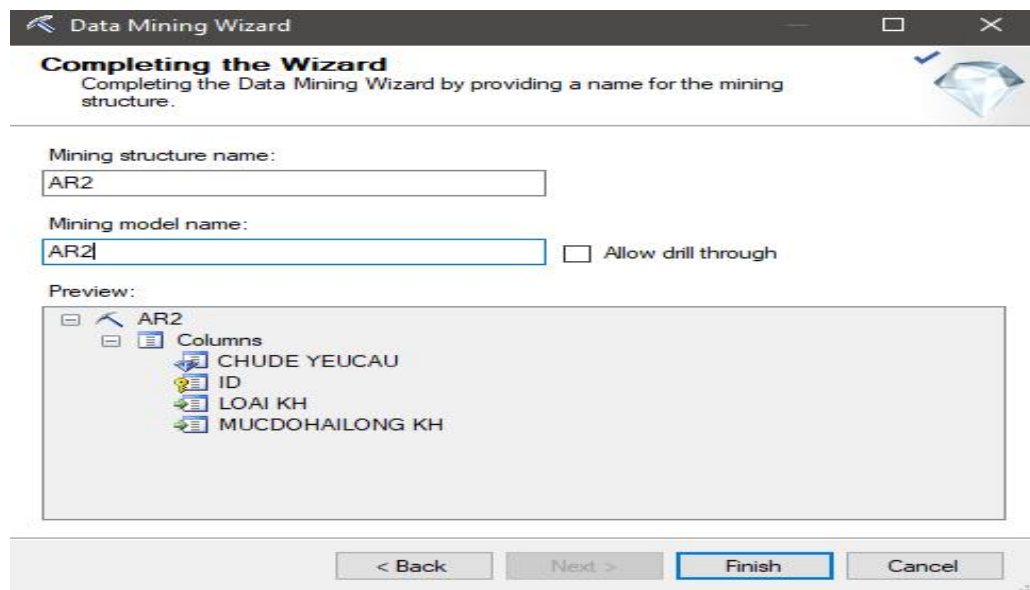
- Chọn thuộc tính để Training:

Chọn các thuộc tính Input: MUCDOHAILONG\_KH , LOAI\_KH, thuộc tính CHUDE\_YEUCAU chọn Predictable, ignore thuộc tính KENH\_LIENLAC, QUAN.



**Hình 3.26 Chọn thuộc tính để Training**

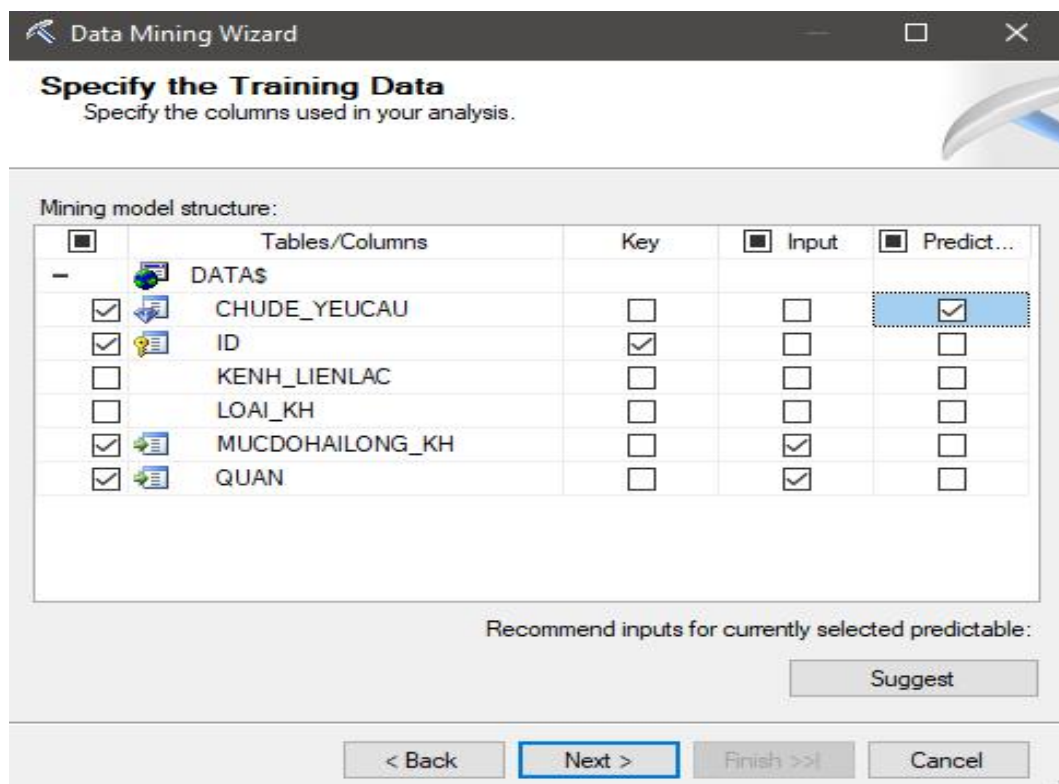
- Đặt tên cho mô hình này là AR2 và chọn Finish.



**Hình 3.27 Đặt tên mô hình luật kết hợp.**

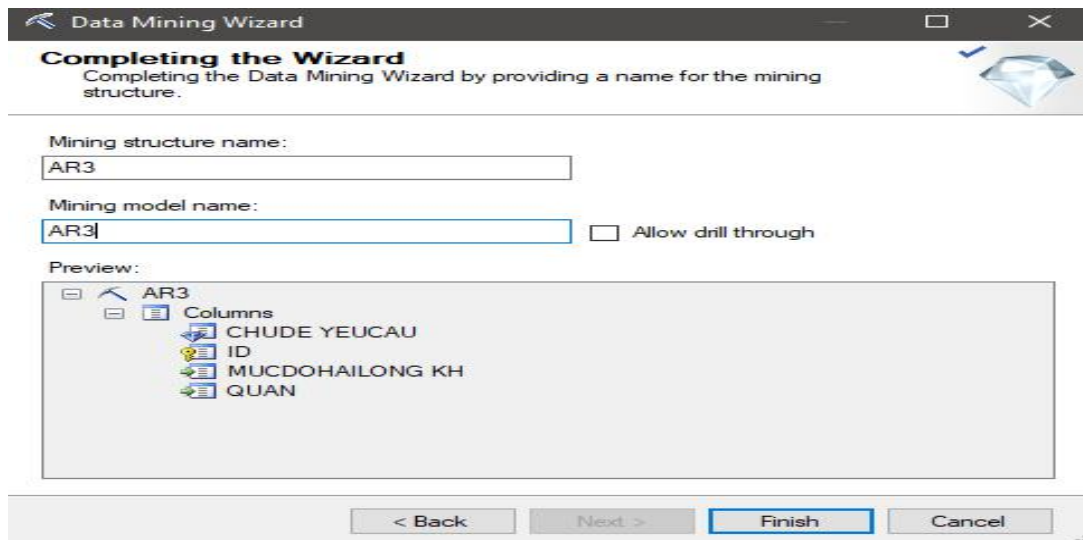
- Chọn thuộc tính để Training:

Chọn các thuộc tính Input: MUCDOHAILONG\_KH, QUAN, thuộc tính CHUDE\_YEUCAU chọn Predictable, ignore thuộc tính KENH\_LIENLAC, LOAI\_KH.



**Hình 3.28 Chọn thuộc tính để Training.**

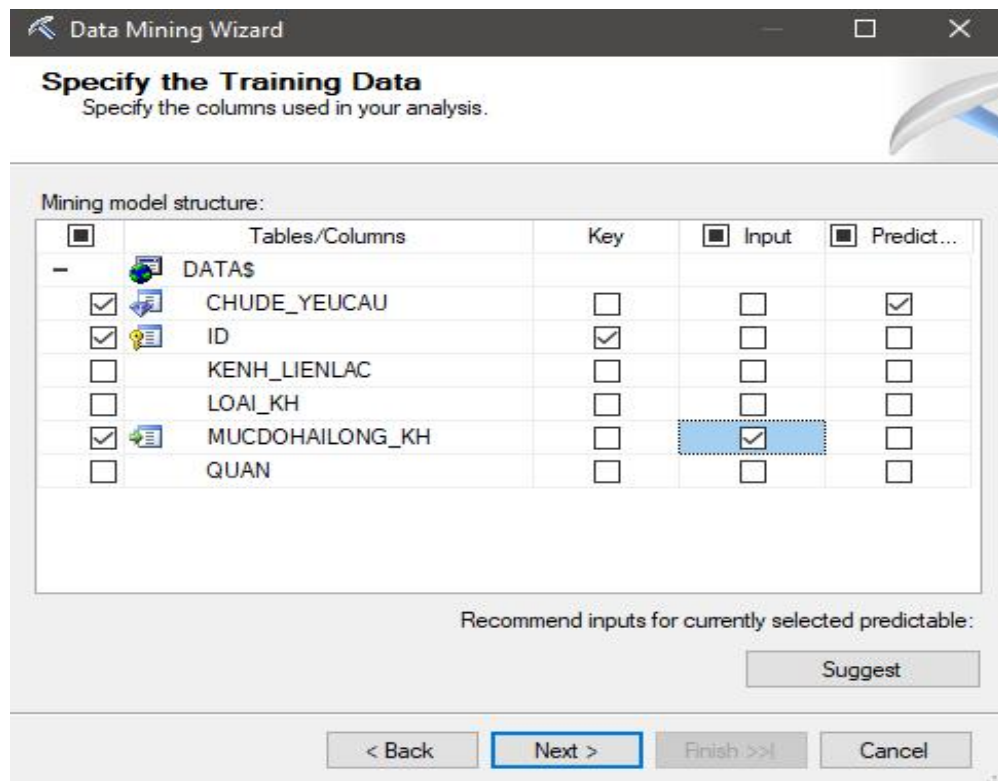
- Đặt tên cho mô hình này là AR3 và chọn Finish.



**Hình 3.29 Đặt tên mô hình luật kết hợp.**

- Chọn thuộc tính để Training:

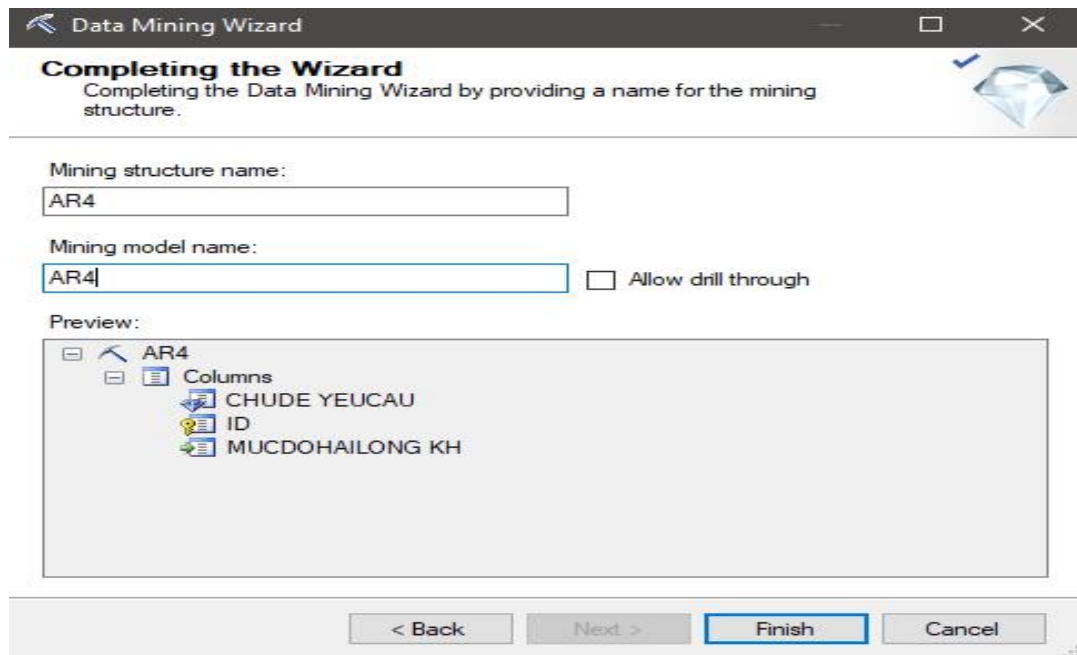
Chọn các thuộc tính Input: MUCDOHAILONG\_KH, thuộc tính CHUDE\_YEUCAU chọn Predictable, ignore thuộc tính KENH\_LIENLAC, LOAI\_KH.



**Hình 3.30 Chọn thuộc tính để Training.**

- Đặt tên cho mô hình này là AR4 và chọn Finish.





**Hình 3.31 Đặt tên mô hình luật kết hợp.**

### ***Thiết lập tham số***

Thuật toán luật kết hợp của Microsoft hỗ trợ một số tham số ảnh hưởng đến hiệu suất và độ chính xác của kết quả mô hình khai thác.

#### **- MAXIMUM\_ITEMSET\_COUNT**

Chỉ số lượng tối đa của các tập phổ biến được sinh ra. Nếu không được quy định số lượng cụ thể, giá trị mặc định sẽ được sử dụng

Mặc định là 200000.

#### **- MAXIMUM\_ITEMSET\_SIZE**

Chỉ định số lượng tối đa các mục được phép trong tập phổ biến. Thiết lập giá trị này là 0 xác định rằng không có giới hạn kích thước của tập phổ biến.

Giá trị này có khả năng có thể giảm thời gian là cần thiết cho việc tạo ra các mô hình, bởi vì việc xử lý của mô hình dừng lại khi giới hạn được đạt đến.

Mặc định là 3.

#### **- MAXIMUM\_SUPPORT**

Chỉ định số lượng tối đa của các trường hợp mà tập phổ biến được hỗ trợ. Tham số này có thể được sử dụng để loại bỏ các mục mà xuất hiện thường xuyên.

Nếu giá trị này nhỏ hơn 1, tức là giá trị đại diện cho một tỷ lệ phần trăm của tổng số các trường hợp. Những giá trị lớn hơn 1 đại diện cho số lượng tuyệt đối của các trường hợp mà có thể chứa các tập phổ biến.

Mặc định là 1.

- MINIMUM\_ITEMSET\_SIZE

Chỉ định số lượng tối thiểu các mục được phép trong tập phổ biến. Nếu tăng giá trị tham số này, mô hình có thể chứa các tập phổ biến ít hơn.

Mặc định là 1.

- MINIMUM\_PROBABILITY

Xác định xác suất tối thiểu mà một luật sinh ra là đúng.

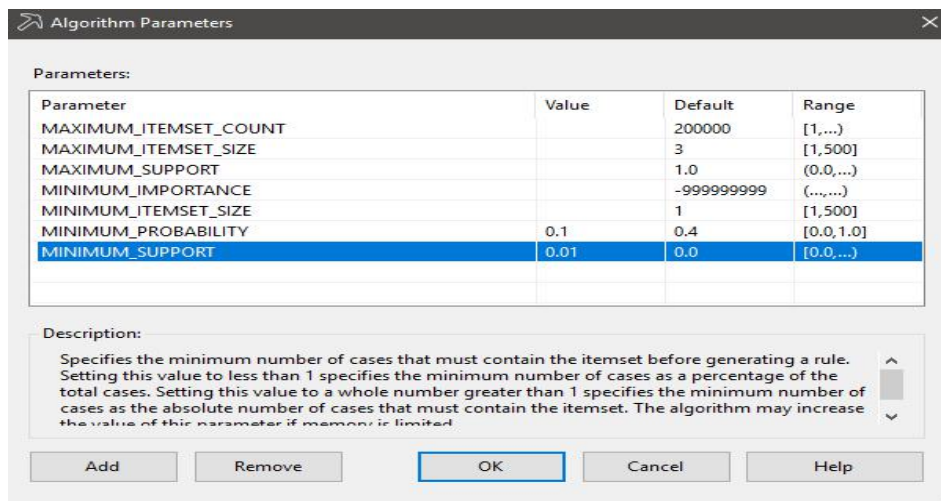
Mặc định là 0.4.

- MINIMUM\_SUPPORT

Chỉ định số lượng tối thiểu các trường hợp mà phải chứa các tập phổ biến trước khi thuật toán tạo ra các tập luật.

Nếu thiết đặt giá trị này nhỏ hơn 1, số lượng tối thiểu các trường hợp được tính như là một tỷ lệ phần trăm của tổng số các trường hợp. Nếu thiết đặt giá trị này là một số nguyên lớn hơn 1, xác định số lượng tối thiểu các trường hợp được tính như số các trường hợp mà phải có các tập phổ biến. Thuật toán có lẽ tự động tăng giá trị của các tham số này nếu bộ nhớ được giới hạn.

Mặc định là 0.0.



**Hình 3.32 Thiết lập tham số mô hình luật kết hợp.**

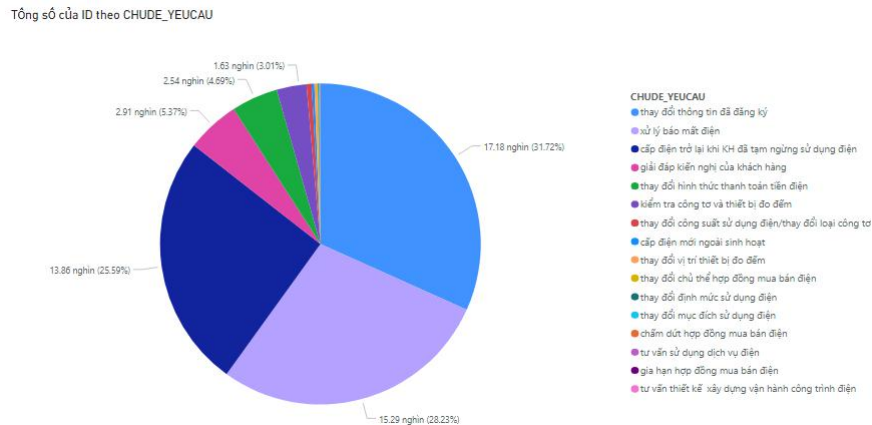


## CHƯƠNG 4

### KẾT QUẢ PHÂN TÍCH

#### 4.1. Kết quả báo cáo BI

##### *Thống kê khách hàng theo từng loại chủ đề yêu cầu*



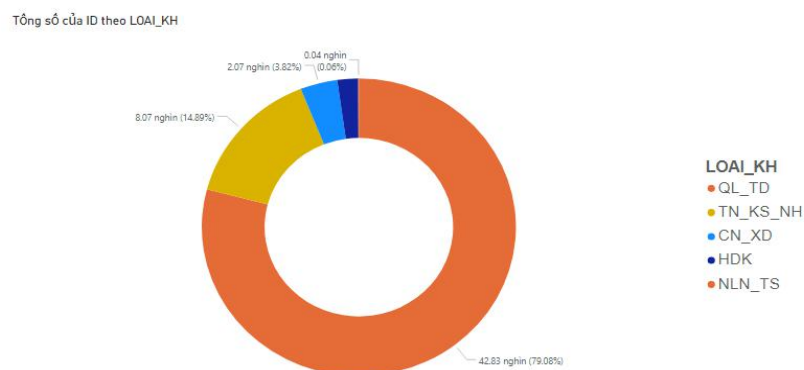
**Hình 4.1. Thống kê khách hàng theo từng loại chủ đề yêu cầu.**

- Nhìn tổng thể biểu đồ trên ta có thể thấy tỷ lệ khách hàng sử dụng dịch vụ thay đổi thông tin đã đăng ký, xử lý báo mất điện, cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện chiếm khác cao và vượt xa so với những dịch vụ còn lại.

##### **Về chủ đề yêu cầu**

- Thay đổi thông tin đã đăng ký: 31,72%.
- Xử lý báo mất điện: 28,23%.
- Cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện: 25,59%.

##### *Thống kê khách hàng theo từng loại khách hàng*



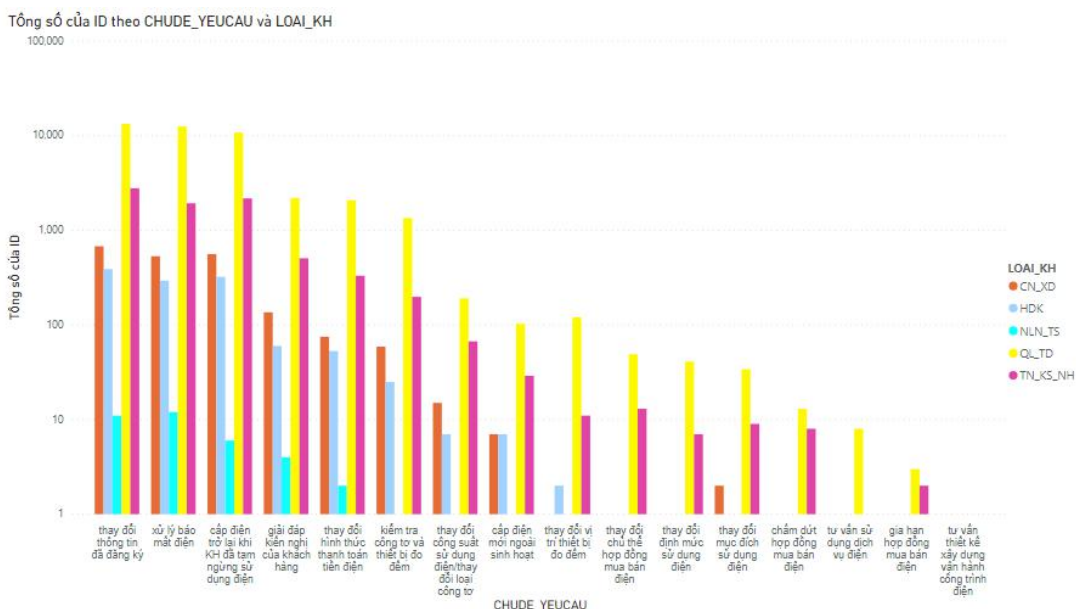
### Hình 4.2 Thống kê khách hàng theo từng loại khách hàng.

- Nhìn tổng thể trên 5 loại khách hàng thì loại quản lý – tiêu dùng chiếm tỷ lệ khá cao là 79,08%.

#### Về loại khách hàng

- Loại khách hàng Thương nghiệp – khách sạn – nhà hàng: 14,89%.
- Loại khách hàng Công nghiệp – xây dựng: 3,82%.
- Loại khách hàng hoạt động khác: 2,14%.
- Loại khách hàng Nông lâm nghiệp – thủy sản: 0,06%.

### Thống kê khách hàng theo từng loại khách hàng và chủ đề yêu cầu



### Hình 4.3 Thống kê khách hàng theo từng loại khách hàng và chủ đề yêu cầu.

#### Về chủ đề yêu cầu

- Nhìn tổng thể ta thấy 3 chủ đề yêu cầu: thay đổi thông tin đã đăng ký, xử lý báo mất điện và cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện được khách hàng yêu cầu nhiều nhất.
- Tiếp đến là chủ đề giải đáp kiến nghị khách hàng và thay đổi hình thức thanh toán tiền điện. Các chủ đề này được yêu cầu bởi tất cả các loại khách hàng.

- Những chủ đề còn lại phần lớn chỉ được yêu cầu bởi 3 đến 4 loại khách hàng. Riêng chủ đề tư vấn thiết kế xây dựng vận hành công trình điện là chỉ có 1 yêu cầu từ loại khách hàng Quản lý - tiêu dùng.

### **Về loại khách hàng**

- Khách hàng thuộc loại Công nghiệp – xây dựng phản hồi hầu hết các chủ đề yêu cầu, trong đó chủ đề thay đổi thông tin đã đăng ký chiếm tỉ lệ cao nhất với 678 khách hàng. Các chủ đề ít được khách hàng phản hồi nhất là:

Chủ đề thay đổi vị trí thiết bị đo đếm: 1 khách hàng.

Chủ đề thay đổi chủ thể hợp đồng mua bán điện: 1 khách hàng.

Chủ đề thay đổi định mức sử dụng điện: 1 khách hàng.

Chủ đề thay đổi mục đích sử dụng điện: 2 khách hàng.

Chủ đề tư vấn sử dụng dịch vụ điện: 1 khách hàng.

- Khách hàng thuộc loại hoạt động khác phần lớn phản hồi về chủ đề thay đổi thông tin đã đăng ký với 389 khách hàng. Các chủ đề ít được khách hàng phản hồi nhất được thống kê như sau:

Chủ đề thay đổi công suất sử dụng điện/thay đổi loại công tơ: 7 khách hàng.

Chủ đề cấp điện mới ngoài sinh hoạt: 7 khách hàng.

Chủ đề thay đổi vị trí thiết bị đo đếm: 2 khách hàng.

Chủ đề tư vấn sử dụng dịch vụ điện: 1 khách hàng.

- Khách hàng thuộc loại Nông lâm nghiệp – thủy sản phản hồi về ít chủ đề nhất, được liệt kê như sau:

Chủ đề thay đổi thông tin đã đăng ký: 11 khách hàng.

Chủ đề xử lý báo mất điện: 12 khách hàng.

Chủ đề cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện: 6 khách hàng.

Chủ đề giải đáp kiến nghị của khách hàng: 4 khách hàng.

Chủ đề thay đổi hình thức thanh toán tiền điện: 2 khách hàng.

- Khách hàng thuộc loại Quản lý – tiêu dùng xuất hiện trong tất cả các chủ đề yêu cầu, tuy nhiên số lượng phản hồi về chủ đề thay đổi thông tin đã đăng ký chiếm tỉ lệ cao nhất với 13323 khách hàng. Những chủ đề ít được khách hàng phản hồi nhất được thống kê như sau:

Chủ đề tư vấn sử dụng dịch vụ điện: 8 khách hàng.

Chủ đề gia hạn hợp đồng mua bán điện: 3 khách hàng

Chủ đề tư vấn thiết kế xây dựng vận hàng công trình điện: 1 khách hàng.

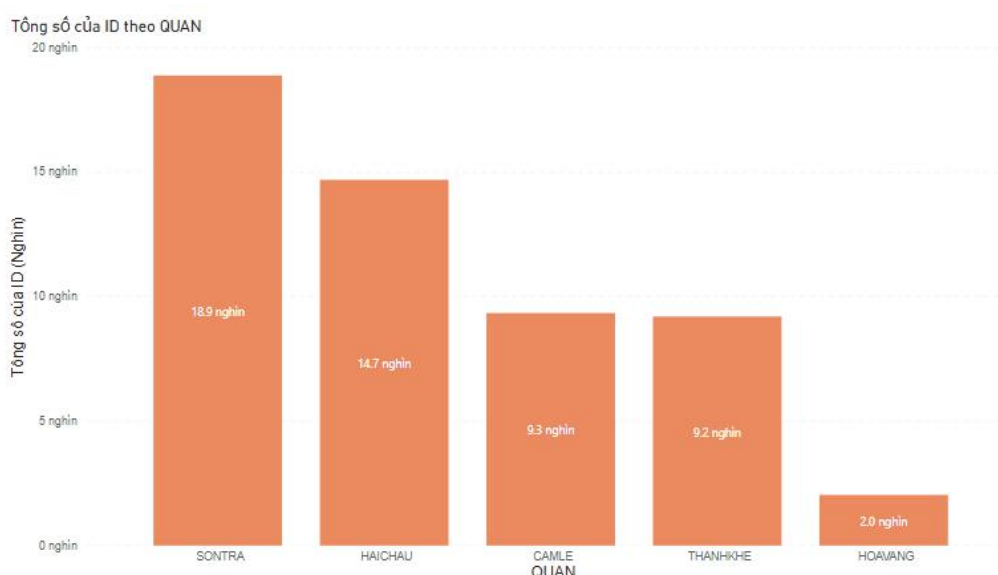
• Khách hàng thuộc Thương nghiệp – khách sạn – nhà hàng có số lượng khách hàng phản hồi tương đối nhưng cao nhất vẫn là chủ đề thay đổi thông tin đã đăng ký với 2780 khách hàng. Các chủ đề còn lại được yêu cầu tương đối, trong đó có 3 chủ đề ít được khách hàng yêu cầu nhất là:

Chủ đề thay đổi định mức sử dụng điện: 7 khách hàng.

Chủ đề tư vấn sử dụng dịch vụ điện: 1 khách hàng.

Chủ đề gia hạn hợp đồng mua bán điện: 2 khách hàng.

### ***Thống kê khách hàng theo từng quận***



***Hình 4.4. Thống kê khách hàng theo từng quận.***

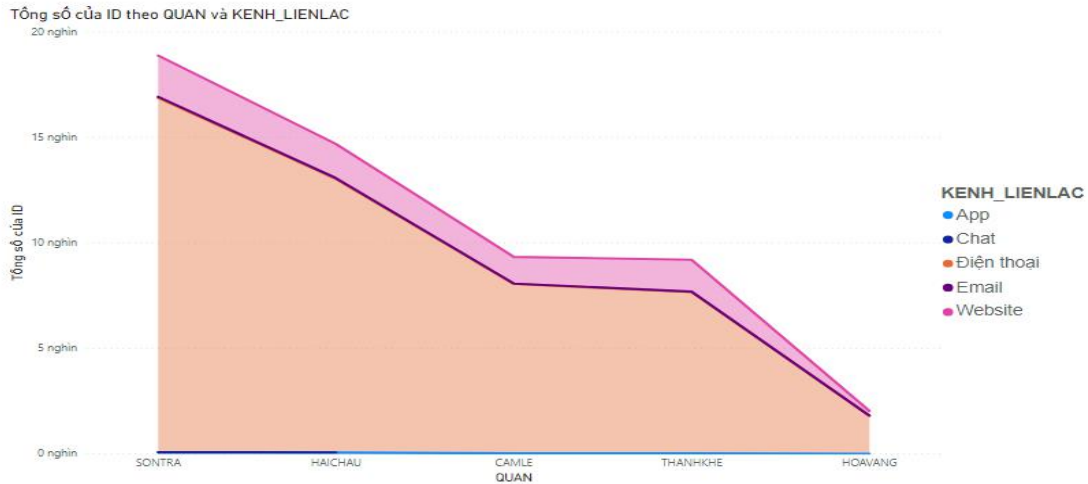
- Nhìn chung ta thấy số lượng khách hàng ở quận Sơn Trà phản hồi nhiều nhất và số lượng phản hồi ít nhất nằm ở quận Hòa Vang.

### **Về số lượng phản hồi của từng quận**

- Quận Sơn Trà: 18886 phản hồi.
- Quận Hải Châu: 14695 phản hồi.
- Quận Cẩm Lệ: 9342 phản hồi.

- Quận Thanh Khê: 9204 phản hồi.
- Quận Hòa Vang: 2037 phản hồi.

### ***Thống kê khách hàng theo từng quận và kênh liên lạc***



***Hình 4.5. Thống kê khách hàng theo từng quận và kênh liên lạc.***

### **Về kênh liên lạc**

- Khách hàng thuộc tất cả các quận hầu hết phản hồi qua kênh liên lạc là điện thoại.
- Khách hàng ít phản hồi qua kênh liên lạc là Email, Chat và App.
- Khách hàng phản hồi qua Website chiếm tương đối.

### **Về các quận**

- Quận Sơn Trà: Nhìn tổng thể ta thấy quận Sơn Trà với số lượng khách hàng phản hồi nhiều nhất, đặc biệt là thông qua kênh liên lạc điện thoại lên đến 16802 khách hàng. Kênh liên lạc Website, App và Email được sử dụng tương đối.

Kênh liên lạc Chat được ít khách hàng sử dụng nhất: 1 khách hàng.

- Quận Hải Châu: Số lượng khách hàng phản hồi thông qua kênh liên lạc điện thoại ở quận Hải Châu chỉ thấp hơn quận Sơn Trà với 12970 khách hàng. Ở quận này, kênh liên lạc Chat chỉ được 1 khách hàng sử dụng, chiếm tỉ lệ thấp nhất.

- Quận Cẩm Lệ: Khách hàng ở quận Cẩm Lệ phản hồi thông qua kênh liên lạc điện thoại với 803 khách hàng. Quận Cẩm Lệ không có khách hàng sử dụng kênh liên lạc Chat.

Kênh liên lạc ít được sử dụng nhất là Email với 15 khách hàng.

- Quận Thanh Khê: Chủ yếu khách hàng ở quận Thanh Khê liên lạc thông qua kênh liên lạc điện thoại với 7647 khách hàng.

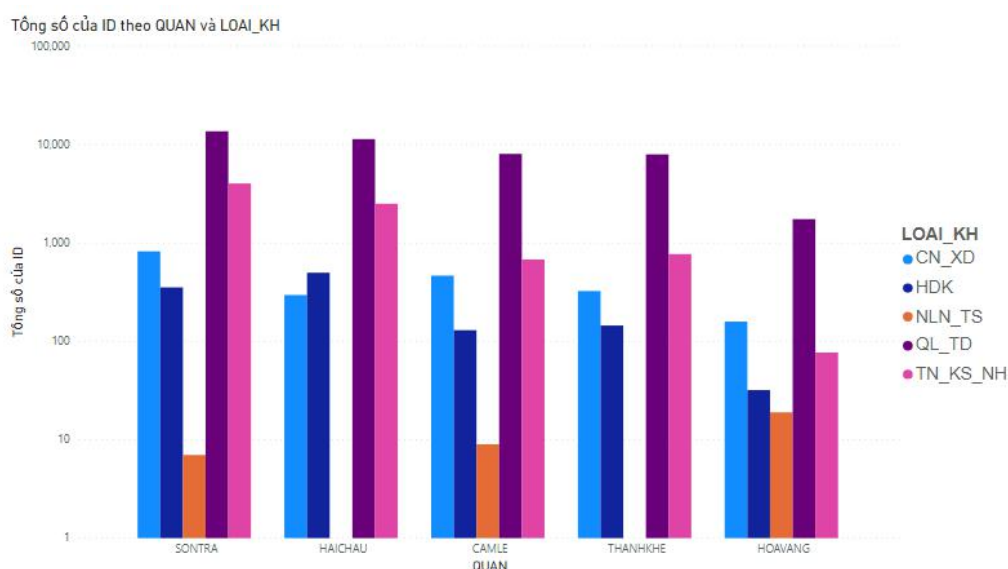
Kênh liên lạc Email ít được khách hàng sử dụng nhất với 22 khách hàng.

- Quận Hòa Vang: Tổng số khách hàng phản hồi thông qua các kênh liên lạc ở quận Hòa Vang ít nhất trong 5 quận, tuy nhiên kênh liên lạc chủ yếu vẫn là điện thoại với 1808 khách hàng. Kênh liên lạc Website chiếm tương đối trong khi kênh liên lạc App và Email đều ít được khách hàng sử dụng:

Kênh liên lạc App: 5 khách hàng.

Kênh liên lạc Email: 2 khách hàng.

### ***Thống kê khách hàng theo từng quận và loại khách hàng***



***Hình 4.6. Thống kê khách hàng theo từng quận và loại khách hàng.***

### **Về loại khách hàng**

- Khách hàng ở tất cả các quận hầu hết là loại khách hàng Quản lý – tiêu dùng.
- Khách hàng thuộc loại Nông lâm nghiệp – thủy sản ít phản hồi nhất.
- Khách hàng thuộc loại Thương nghiệp – khách sạn – nhà hàng, Công nghiệp – xây dựng, hoạt động khác chiếm tương đối.

### **Về các quận**

- Quận Sơn Trà: Số lượng khách hàng phản hồi nhiều nhất ở quận Sơn Trà chủ yếu là loại Quản lý – tiêu dùng với 13673 khách hàng.

Với số lượng 7 khách hàng, loại khách hàng Nông lâm nghiệp – thủy sản chiếm tỉ lệ thấp nhất.

- Quận Hải Châu: Số lượng khách hàng phản hồi ở quận Hải Châu cao thứ 2 trong tất cả các quận, các khách hàng này hầu hết là loại Quản lý – tiêu dùng với 11392 khách hàng.

Loại khách hàng Công nghiệp – xây dựng chiếm tỉ lệ thấp nhất với 296 khách hàng.

- Quận Cẩm Lệ: Các khách hàng ở quận Cẩm Lệ thường là loại khách hàng Quản lý - tiêu dùng với 8057 khách hàng.

Với số lượng là 9 khách hàng, loại khách hàng Nông lâm nghiệp – thủy sản ở quận Cẩm Lệ chiếm tỉ lệ thấp nhất.

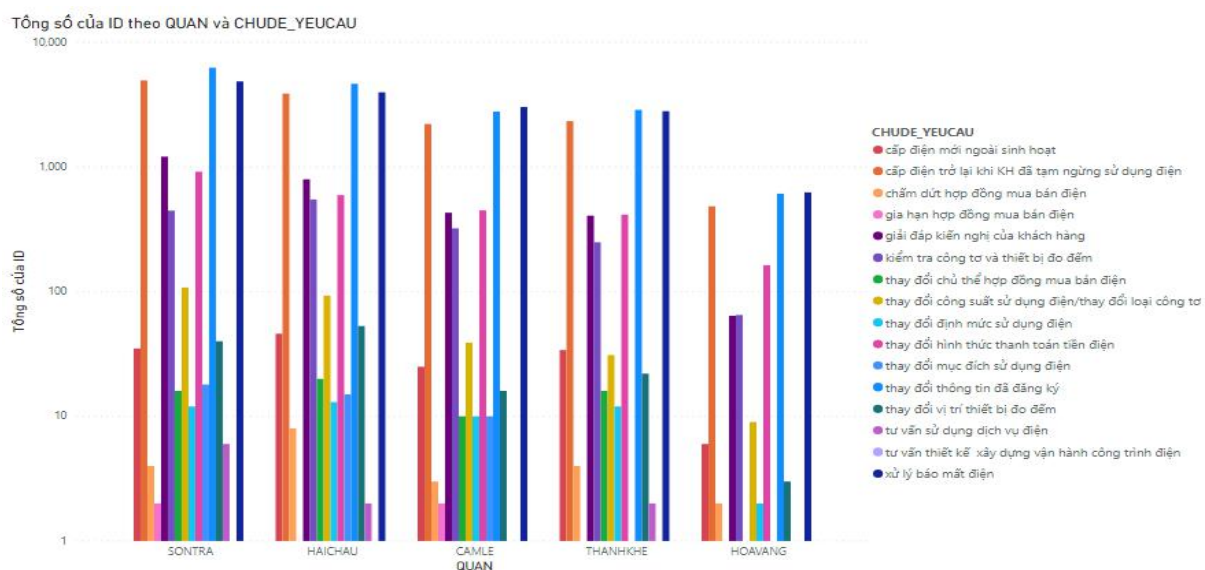
- Quận Thanh Khê: Chủ yếu khách hàng ở quận Thanh Khê là loại khách hàng Quản lý – tiêu dùng với 7962 khách hàng.

Loại khách hàng Hoạt động khác ở quận Thanh Khê có số lượng khách hàng phản hồi nhất với 145 khách hàng.

- Quận Hòa Vang: Khách hàng ở quận Hòa Vang có số lượng phản hồi ít nhất trong số 5 quận, nhưng hầu hết vẫn là loại khách hàng Quản lý – tiêu dùng với 1750 khách hàng.

Loại khách hàng Nông lâm nghiệp – thủy sản với 19 khách hàng chính là loại khách hàng ít phản hồi nhất, có tỉ lệ thấp hơn hẳn so với các loại khách hàng còn lại.

### ***Thống kê khách hàng theo từng quận và chủ đề yêu cầu***



**Hình 4.7. Thống kê khách hàng theo từng quận và chủ đề yêu cầu.**

### Về chủ đề yêu cầu

- Khách hàng ở tất cả các quận phần lớn phản hồi về 3 chủ đề: cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện, thay đổi thông tin đã đăng ký và xử lý báo mất điện.

- Khách hàng rất ít phản hồi về các chủ đề: cấp điện mới ngoài sinh hoạt, chấm dứt hợp đồng mua bán điện, gia hạn hợp đồng mua bán điện, thay đổi chủ thẻ hợp đồng mua bán điện, thay đổi định mức sử dụng điện, thay đổi mục đích sử dụng điện, thay đổi vị trí thiết bị đo đếm, tư vấn sử dụng dịch vụ điện và tư vấn thiết kế xây dựng vận hành công trình điện.

- Số lượng khách hàng phản hồi về 4 chủ đề: giải đáp kiến nghị khách hàng, kiểm tra công tơ và thiết bị đo đếm, thay đổi công suất sử dụng điện/thay đổi loại công tơ, thay đổi hình thức thanh toán tiền điện chiếm tương đối.

### Về các quận

- Quận Sơn Trà: Khách hàng ở quận Sơn Trà có phản hồi về chủ đề thay đổi thông tin đã đăng ký cao nhất với 6261 khách hàng.

Chủ đề cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện được 4949 khách hàng phản hồi, đứng thứ 2 trong số các chủ đề yêu cầu.

Chủ đề được ít khách hàng phản hồi nhất là chủ đề gia hạn hợp đồng mua bán điện với 2 khách hàng.



- Quận Hải Châu: Khách hàng ở quận Hải Châu chủ yếu phản hồi về chủ đề thay đổi thông tin đã đăng ký với 4652 khách hàng.

Hai chủ đề ít được khách hàng phản hồi nhất là:

Chủ đề chấm dứt hợp đồng mua bán điện: 8 khách hàng.

Chủ đề tư vấn sử dụng dịch vụ điện: 2 khách hàng.

- Quận Cẩm Lệ: Các khách hàng ở quận Cẩm Lệ phản hồi về chủ đề thay đổi thông tin đã đăng ký nhiều nhất với 2785 khách hàng.

Các chủ đề ít được khách hàng yêu cầu nhất là:

Chủ đề chấm dứt hợp đồng mua bán điện: 3 khách hàng.

Chủ đề gia hạn hợp đồng mua bán điện: 2 khách hàng.

Chủ đề tư vấn sử dụng dịch vụ điện: 1 khách hàng.

- Quận Thanh Khê: Hầu hết khách hàng ở quận Thanh Khê phản hồi về chủ đề thay đổi thông tin đã đăng ký với 2872 khách hàng.

Những chủ đề khác chiếm tương đối, ngoại trừ 5 chủ đề sau được ít khách hàng phản hồi nhất:

Chủ đề chấm dứt hợp đồng mua bán điện: 4 khách hàng.

Chủ đề gia hạn hợp đồng mua bán điện: 1 khách hàng.

Chủ đề thay đổi mục đích sử dụng điện: 1 khách hàng.

Chủ đề tư vấn sử dụng dịch vụ điện: 2 khách hàng.

Chủ đề tư vấn thiết kế xây dựng vận hành công trình điện: 1 khách hàng.

- Quận Hòa Vang: Quận Hòa Vang có số lượng khách hàng phản hồi ít nhất trong 5 quận thuộc thành phố Đà Nẵng, tuy nhiên các khách hàng ở quận Hòa Vang chủ yếu phản hồi về chủ đề thay đổi thông tin đã đăng ký với 611 khách hàng.

Trong khi các chủ đề khác chiếm tương đối thì những chủ đề sau đây chiếm tỉ lệ thấp hơn hẳn, với số lượng khách hàng phản hồi là:

Chủ đề cấp điện mới ngoài sinh hoạt: 6 khách hàng.

Chủ đề chấm dứt hợp đồng mua bán điện: 2 khách hàng.

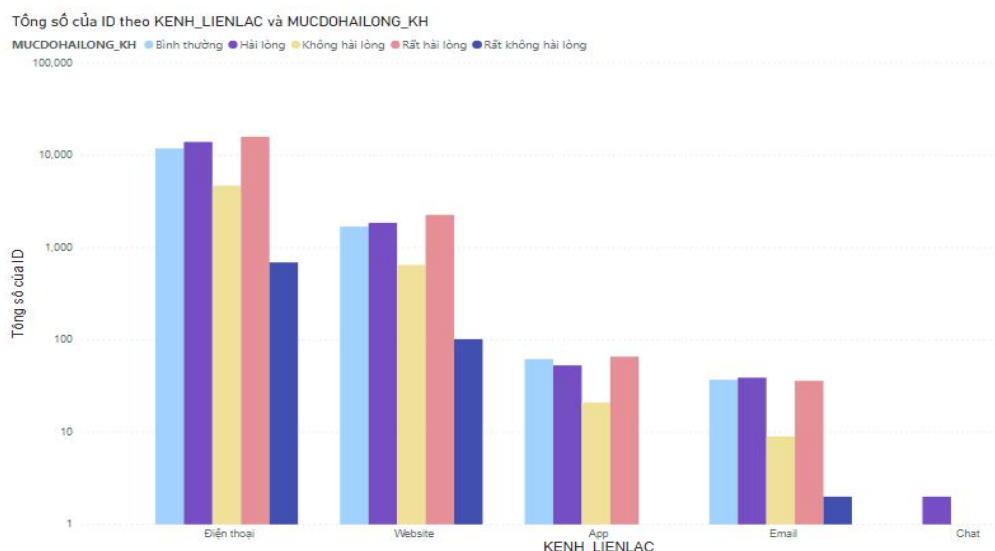
Chủ đề thay đổi chủ thể hợp đồng mua bán điện: 1 khách hàng.

Chủ đề thay đổi định mức sử dụng điện: 2 khách hàng.

Chủ đề thay đổi mục đích sử dụng điện: 1 khách hàng.

Chủ đề thay đổi vị trí thiết bị đo đếm: 3 khách hàng.

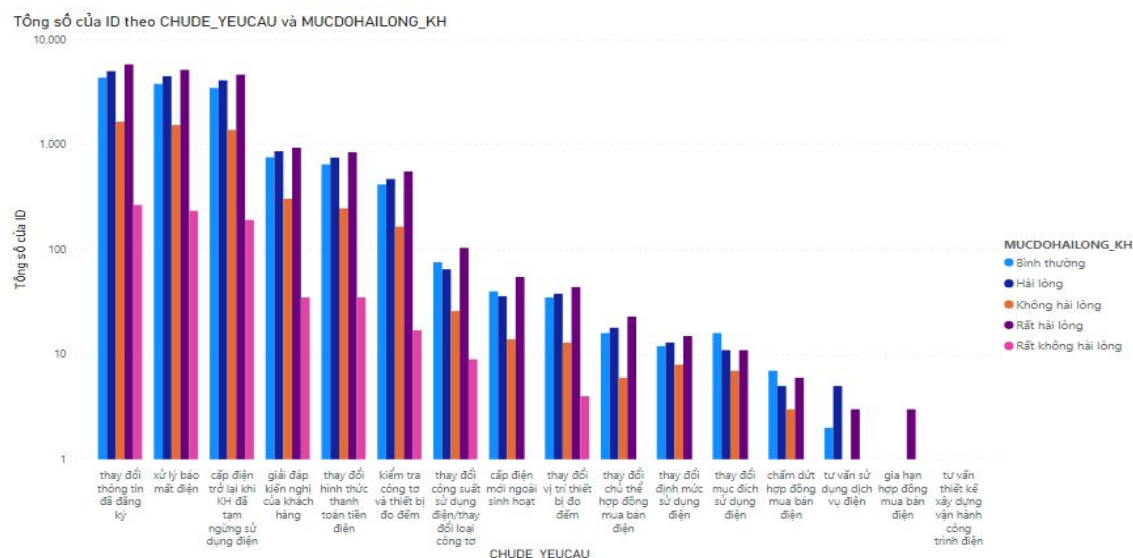
## Thống kê mức độ hài lòng khách hàng theo kênh liên lạc



**Hình 4.8. Thống kê mức độ hài lòng khách hàng theo kênh liên lạc.**

- Nhìn chung, kênh liên lạc Điện thoại được khách hàng sử dụng nhiều nhất và cũng chính là kênh liên lạc có số lượng khách hàng đánh giá không hài lòng cao nhất: 5409 khách hàng. Ngược lại, kênh liên lạc Website có số lượng khách hàng sử dụng ít hơn nhưng lại là kênh liên lạc có mức độ hài lòng khách hàng cao nhất: 4133 khách hàng.

## Thống kê khách hàng theo mức độ hài lòng khách hàng và chủ đề yêu cầu



**Hình 4.9. Thống kê khách hàng theo mức độ hài lòng khách hàng và chủ đề yêu cầu.**

- Nhìn tổng thể, các chủ đề được yêu cầu nhiều nhất là: thay đổi thông tin đã đăng ký, xử lý báo mất điện, cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện.

- Chủ đề thay đổi thông tin đã đăng ký có số lượng khách hàng phản hồi nhiều nhất, được thống kê như sau:

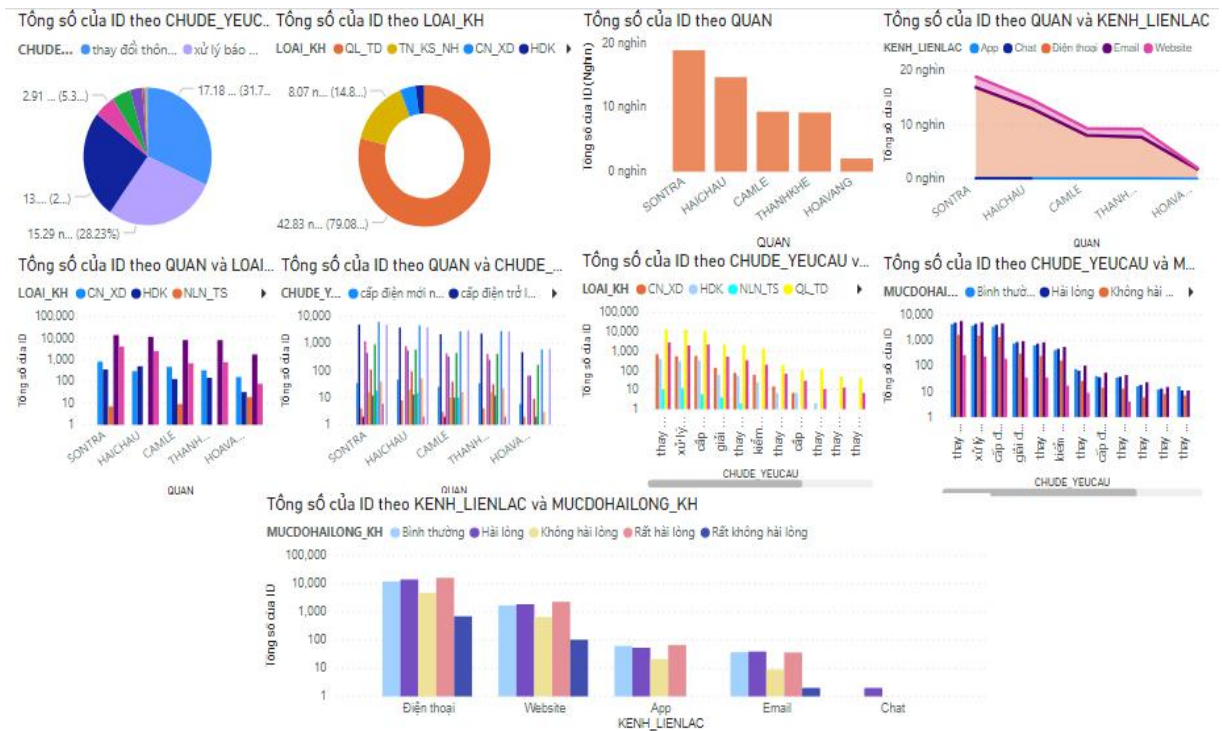
- Mức độ rất hài lòng: 5844 khách hàng.
- Mức độ hài lòng: 5037 khách hàng.
- Mức độ bình thường: 4372 khách hàng.
- Mức độ không hài lòng: 1661 khách hàng.
- Mức độ rất không hài lòng: 267 khách hàng.

- Chủ đề xử lý báo mất điện có số lượng khách hàng phản hồi cao thứ hai, chỉ đứng sau chủ đề thay đổi thông tin đã đăng ký. Độ hài lòng của khách hàng khi phản hồi về chủ đề này được thống kê như sau:

- Mức độ rất hài lòng: 5186 khách hàng.
- Mức độ hài lòng: 4513 khách hàng.
- Mức độ bình thường: 3807 khách hàng.
- Mức độ không hài lòng: 1549 khách hàng.
- Mức độ rất không hài lòng: 235 khách hàng.

- Chủ đề cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện được khách hàng đánh giá mức độ hài lòng khi phản hồi như sau:

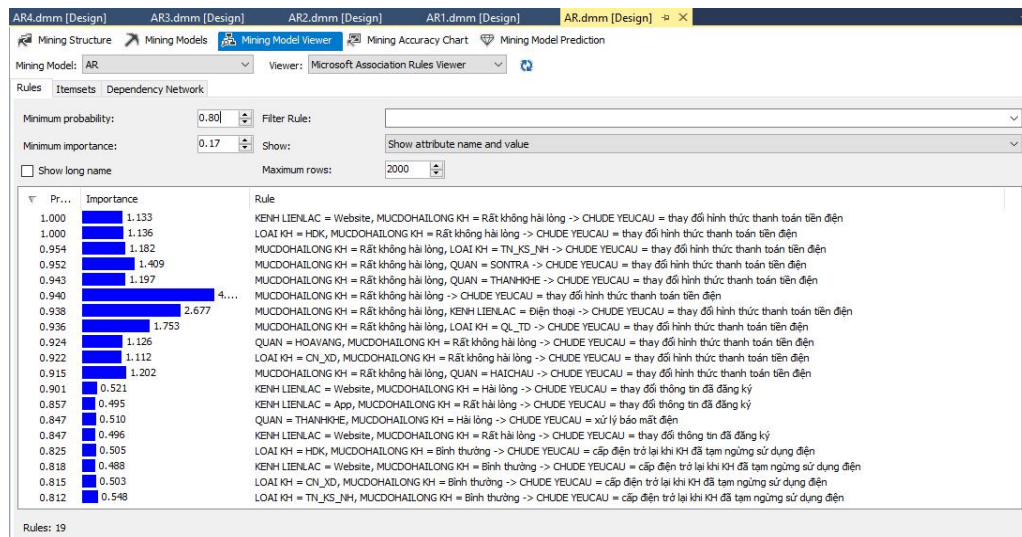
- Mức độ rất hài lòng: 4674 khách hàng.
- Mức độ hài lòng: 4120 khách hàng.
- Mức độ bình thường: 3486 khách hàng.
- Mức độ không hài lòng: 1387 khách hàng.
- Mức độ rất không hài lòng: 192 khách hàng.



**Hình 4.10. Dashboard thống kê số lượng khách hàng tổng hợp.**

## 4.2. Kết quả mô hình luật kết hợp

### Kết quả mô hình AR



**Hình 4.11. Kết quả mô hình luật kết hợp**

Dựa trên kết quả này, ta có thể thấy, mô hình đã cho ra được 19 luật kết hợp mạnh nhất như sau:

- **Luật 1:** Với kênh liên lạc là Website, mức độ hài lòng khách hàng là rất không hài lòng thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 100%.

- **Luật 2:** Với loại khách hàng là HDK, mức độ hài lòng khách hàng là rất không hài lòng thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 100%.

- **Luật 3:** Với mức độ hài lòng khách hàng là rất không hài lòng, loại khách hàng là thương nghiệp - khách sạn - nhà hàng thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 95,4%

- **Luật 4:** Với mức độ hài lòng khách hàng là rất không hài lòng, quận Sơn Trà thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 95,2%

- **Luật 5:** Với mức độ hài lòng khách hàng là rất không hài lòng, quận Thanh Khê thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 94,3%

- **Luật 6:** Với mức độ hài lòng khách hàng là rất không hài lòng thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 94%

- **Luật 7:** Với mức độ hài lòng khách hàng là rất không hài lòng, kênh liên lạc là điện thoại thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 93,8%.

- **Luật 8:** Với mức độ hài lòng khách hàng là rất không hài lòng, loại khách hàng là quản lý-tiêu dùng thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 93,6%.

- **Luật 9:** Với quận là Hòa Vang, mức độ hài lòng khách hàng là rất không hài lòng thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 92,4%.

- **Luật 10:** Với loại khách hàng là công nghiệp - xây dựng, mức độ hài lòng khách hàng là rất không hài lòng thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 92,2%.

- **Luật 11:** Với mức độ hài lòng khách hàng là rất không hài lòng, quận là Hải Châu thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 91,5%

- **Luật 12:** Với kênh liên lạc là Website, mức độ hài lòng khách hàng là hài lòng thì chủ đề yêu cầu là thay đổi thông tin đã đăng ký với xác suất 90,1%.

- **Luật 13:** Với kênh liên lạc là App, mức độ hài lòng là rất hài lòng thì chủ đề yêu cầu là thay đổi thông tin đã đăng ký với xác suất 85,7%.

- **Luật 14:** Với quận là Thanh Khê, mức độ hài lòng khách hàng là hài lòng thì chủ đề là xử lý báo mất điện với xác suất 84,7%.

- **Luật 15:** Với kênh liên lạc là Website, mức độ hài lòng khách hàng là rất hài lòng thì chủ đề yêu cầu là thay đổi thông tin đã đăng ký với xác suất là 84,7%.

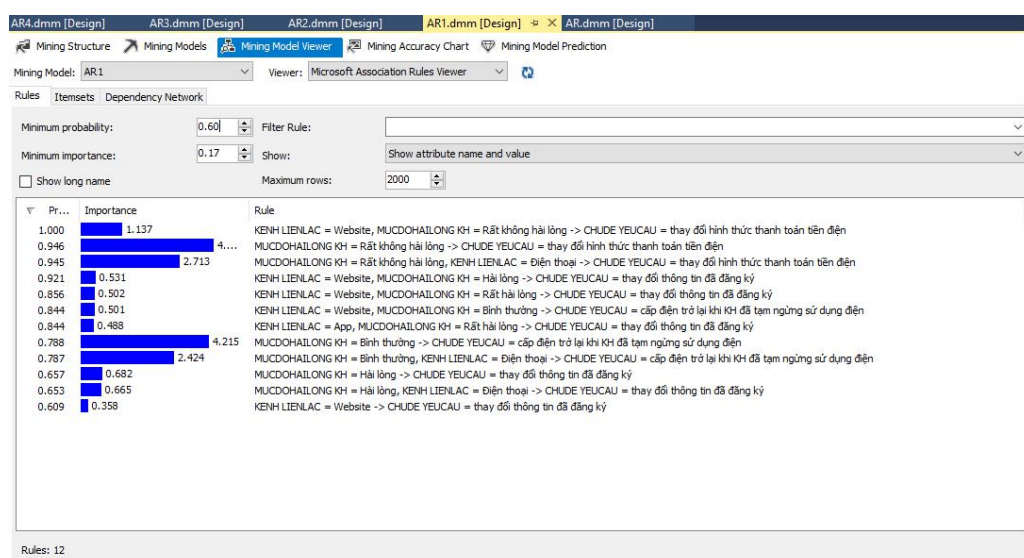
- **Luật 16:** Với loại khách hàng là HDK, mức độ hài lòng khách hàng là bình thường thì chủ đề yêu cầu là cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện với xác suất 82,5%.

- **Luật 17:** Với kênh liên lạc là Website, mức độ hài lòng khách hàng là bình thường thì chủ đề yêu cầu là cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện với xác suất 81,8%.

- **Luật 18:** Với loại khách hàng là công nghiệp - xây dựng, mức độ hài lòng khách hàng là bình thường thì chủ đề yêu cầu là cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện với xác suất 81,5%.

- **Luật 19:** Với loại khách hàng là thương nghiệp - khách sạn - nhà hàng, mức độ hài lòng khách hàng là bình thường thì chủ đề yêu cầu là cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện với xác suất 81,2%.

### Kết quả mô hình AR1



**Hình 4.12. Kết quả mô hình luật kết hợp**

Dựa trên kết quả này, ta có thể thấy, mô hình đã cho ra được 12 luật kết hợp mạnh nhất như sau:

- **Luật 1:** Với kênh liên lạc là Website, mức độ hài lòng khách hàng là rất không hài lòng thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 100%.

- **Luật 2:** Với mức độ hài lòng khách hàng là rất không hài lòng thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 94,6%.

- **Luật 3:** Với mức độ hài lòng khách hàng là rất không hài lòng, kênh liên lạc là điện thoại thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 94,5%.

- **Luật 4:** Với kênh liên lạc là Website, mức độ hài lòng khách hàng là hài lòng thì chủ đề yêu cầu là thay đổi thông tin đã đăng ký với xác suất là 92,1%.

- **Luật 5:** Với kênh liên lạc là Website, mức độ hài lòng khách hàng là rất hài lòng thì chủ đề yêu cầu là thay đổi thông tin đã đăng ký với xác suất 85,6%.

- **Luật 6:** Với kênh liên lạc là App, mức độ hài lòng là rất hài lòng thì chủ đề yêu cầu là thay đổi thông tin đã đăng ký với xác suất 84,4%

- **Luật 7:** Với kênh liên lạc là Website, mức độ hài lòng khách hàng là bình thường thì chủ đề yêu cầu là cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện với xác suất 84,4%.

- **Luật 8:** Với mức độ hài lòng khách hàng là bình thường thì chủ đề yêu cầu là cấp điện trở lại sau khi khách hàng đã tạm ngừng sử dụng điện với xác suất là 78,8%.

- **Luật 9:** Với mức độ hài lòng là bình thường, kênh liên lạc là điện thoại thì chủ đề yêu cầu là cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện với xác suất 78,7%.

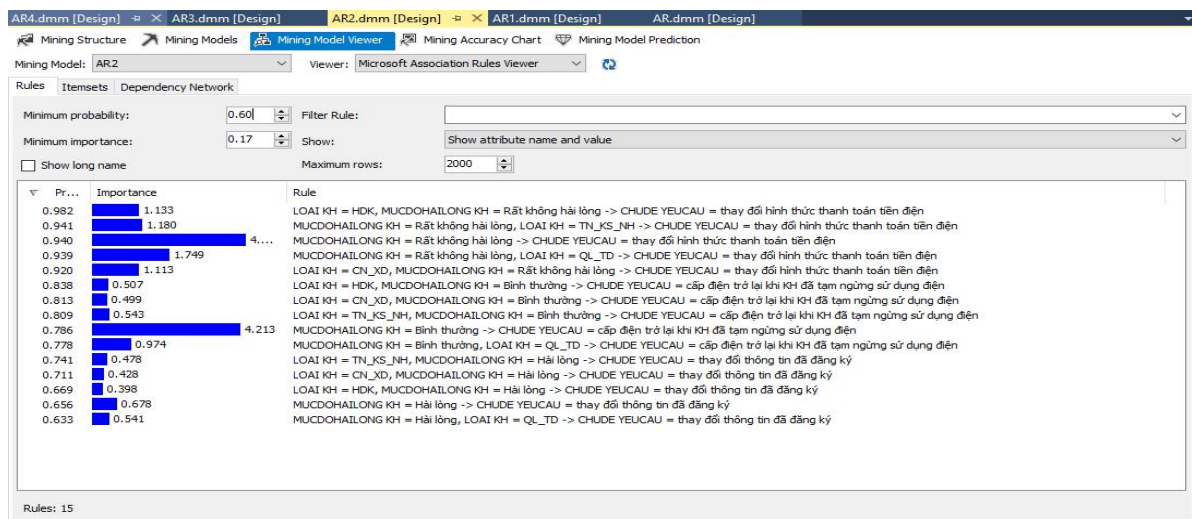
- **Luật 10:** Với mức độ hài lòng khách hàng là hài lòng thì chủ đề yêu cầu là thay đổi thông tin đã đăng ký với xác suất 65,7%.

- **Luật 11:** Với mức độ hài lòng là hài lòng, kênh liên lạc là điện thoại thì chủ đề yêu cầu là thay đổi thông tin đã đăng ký với xác suất 65,3%.

- **Luật 12:** Với kênh liên lạc là Website thì chủ đề yêu cầu là thay đổi thông tin đã đăng ký với xác suất 60,9%.

**Kết quả mô hình AR2 (tiêu đề để cùng trang với nội dung)**





**Hình 4.13. Kết quả mô hình luật kết hợp**

Dựa trên kết quả này, ta có thể thấy, mô hình đã cho ra được 15 luật kết hợp mạnh nhất như sau:

- **Luật 1:** Với loại khách hàng là HDK, mức độ hài lòng khách hàng là rất không hài lòng thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 98,2%.

- **Luật 2:** Với mức độ hài lòng khách hàng là rất không hài lòng, loại khách hàng là thương nghiệp - khách sạn - nhà hàng thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 94,1%.

- **Luật 3:** Với mức độ hài lòng khách hàng là rất không hài lòng thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 94%.

- **Luật 4:** Với mức độ hài lòng khách hàng là rất không hài lòng, loại khách hàng là quản lý - tiêu dùng thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 93,9%.

- **Luật 5:** Với loại khách hàng là công nghiệp - xây dựng, mức độ hài lòng khách hàng là rất không hài lòng thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 92%.

- **Luật 6:** Với loại khách hàng là HDK, mức độ hài lòng khách hàng là bình thường thì chủ đề yêu cầu là cấp điện trở lại sau khi khách hàng đã tạm ngừng sử dụng điện với xác suất 83,8%.



- **Luật 7:** Với loại khách hàng là công nghiệp - xây dựng, mức độ hài lòng khách hàng là bình thường thì chủ đề yêu cầu là cấp điện trở lại sau khi khách hàng đã tạm ngừng sử dụng điện với xác suất 81,3%.

- **Luật 8:** Với loại khách hàng là thương nghiệp - khách sạn - nhà hàng, mức độ hài lòng khách hàng là bình thường thì chủ đề yêu cầu là cấp điện trở lại sau khi khách hàng đã tạm ngừng sử dụng điện với xác suất 80,9%.

- **Luật 9:** Với mức độ hài lòng khách hàng là bình thường thì chủ đề yêu cầu là cấp điện trở lại sau khi khách hàng đã tạm ngừng sử dụng điện với xác suất 78,6%.

- **Luật 10:** Với mức độ hài lòng khách hàng là bình thường, loại khách hàng là quản lý-tiêu dùng thì chủ đề yêu cầu là cấp điện trở lại sau khi khách hàng đã tạm ngừng sử dụng điện với xác suất 77,8%.

- **Luật 11:** Với loại khách hàng là thương nghiệp - khách sạn - nhà hàng, mức độ hài lòng khách hàng là hài lòng thì chủ đề yêu cầu là thay đổi thông tin đã đăng ký với xác suất 74,1%.

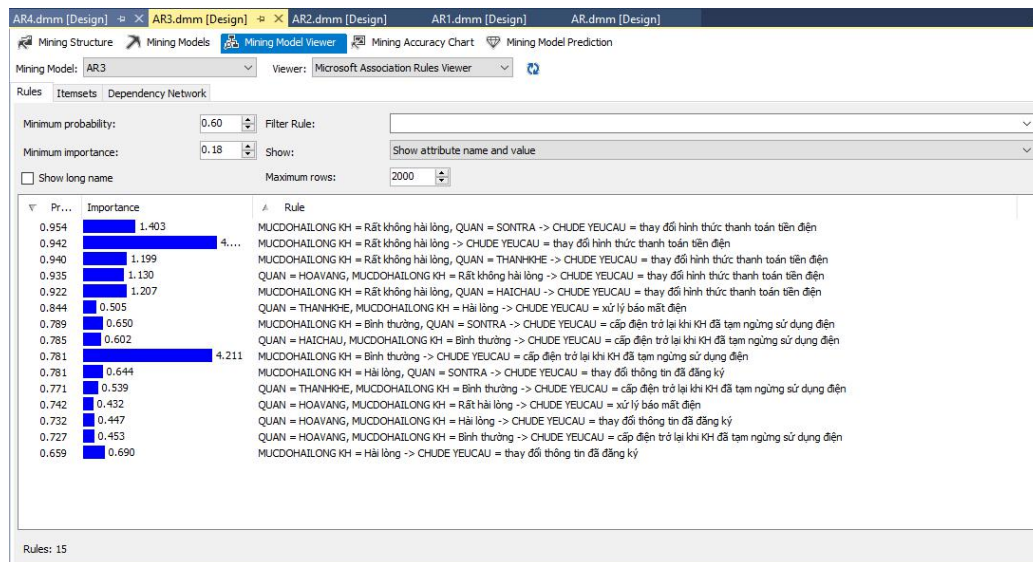
- **Luật 12:** Với loại khách hàng là công nghiệp - xây dựng, mức độ hài lòng khách hàng là hài lòng thì chủ đề yêu cầu là thay đổi thông tin đã đăng ký với xác suất 71,1%.

- **Luật 13:** Với loại khách hàng là HDK, mức độ hài lòng khách hàng là hài lòng thì chủ đề yêu cầu là thay đổi thông tin đã đăng ký với xác suất 66,9%.

- **Luật 14:** Với mức độ hài lòng khách hàng là hài lòng thì chủ đề yêu cầu là thay đổi thông tin đã đăng ký với xác suất 65,6%.

- **Luật 15:** Với mức độ hài lòng khách hàng là hài lòng, loại khách hàng là quản lý-tiêu dùng thì chủ đề yêu cầu là thay đổi thông tin đã đăng ký với xác suất 63,3%.

### ***Kết quả mô hình AR3***



**Hình 4.14. Kết quả mô hình luật kết hợp**

Dựa trên kết quả này, ta có thể thấy, mô hình đã cho ra được 15 luật kết hợp mạnh nhất như sau:

- **Luật 1:** Với mức độ hài lòng khách hàng là rất không hài lòng, quận là Sơn Trà thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 95,4%.
- **Luật 2:** Với mức độ hài lòng khách hàng là rất không hài lòng thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 94,2%.
- **Luật 3:** Với mức độ hài lòng là rất không hài lòng, quận là Thanh Khê thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 94%.
- **Luật 4:** Với quận là Hòa Vang, mức độ hài lòng khách hàng là rất không hài lòng thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 93,5%.
- **Luật 5:** Với mức độ hài lòng khách hàng là rất không hài lòng, quận là Hải Châu thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 92,2%.
- **Luật 6:** Với quận là Thanh Khê, mức độ hài lòng khách hàng là hài lòng thì chủ đề yêu cầu là xử lý báo mất điện với xác suất 84,4%.
- **Luật 7:** Với mức độ hài lòng là bình thường, quận là Sơn Trà thì chủ đề yêu cầu là cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện với xác suất 78,9%.
- **Luật 8:** Với quận là Hải Châu, mức độ hài lòng khách hàng là bình thường thì chủ đề yêu cầu là cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện với xác suất 78,5%.

- **Luật 9:** Với mức độ hài lòng khách hàng là bình thường thì chủ đề yêu cầu là cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện với xác suất 78,1%.

- **Luật 10:** Với mức độ hài lòng khách hàng là hài lòng, quận là Sơn Trà thì chủ đề yêu cầu là thay đổi thông tin đã đăng ký với xác suất 78,1%.

- **Luật 11:** Với quận là Thanh Khê, mức độ hài lòng khách hàng là bình thường thì chủ đề yêu cầu là cấp điện với xác suất 77,1%.

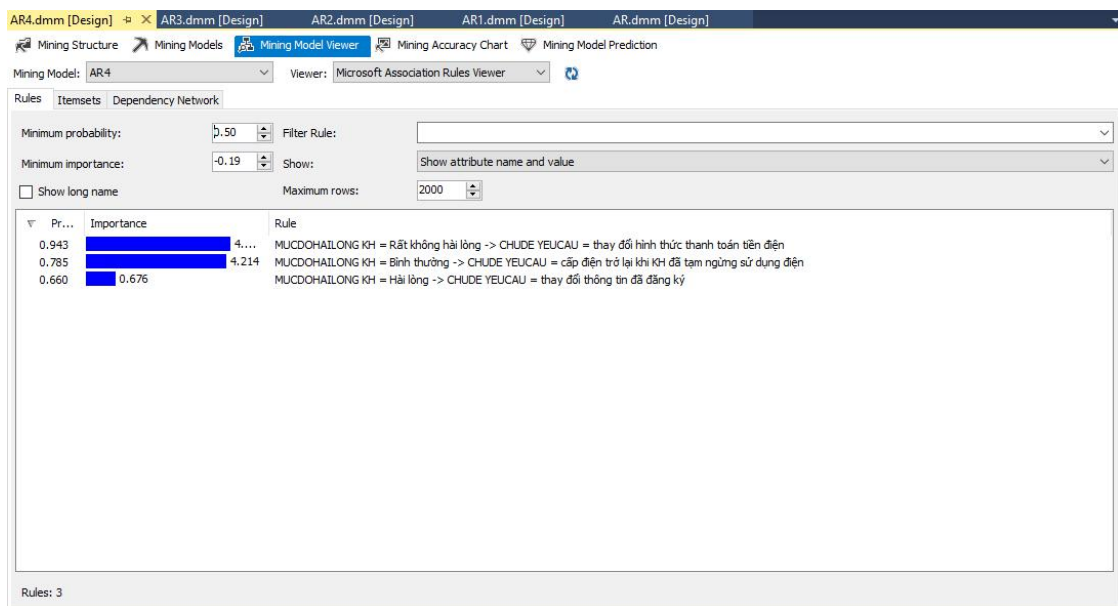
- **Luật 12:** Với quận là Hòa Vang, mức độ hài lòng khách hàng là rất hài lòng thì chủ đề yêu cầu là xử lý báo mất điện với xác suất 74,2%.

- **Luật 13:** Với quận là Hòa Vang, mức độ hài lòng khách hàng là hài lòng thì chủ đề yêu cầu là thay đổi thông tin đã đăng ký với xác suất 73,2%.

- **Luật 14:** Với quận là Hòa Vang, mức độ hài lòng khách hàng là bình thường thì chủ đề yêu cầu là cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện với xác suất 72,7%.

- **Luật 15:** Với mức độ hài lòng khách hàng là hài lòng thì chủ đề yêu cầu là thay đổi thông tin đã đăng ký với xác suất 65,9%.

### Kết quả mô hình AR4



**Hình 4.15. Kết quả mô hình luật kết hợp**

Dựa trên kết quả này, ta có thể thấy, mô hình đã cho ra được 3 luật kết hợp mạnh nhất như sau:

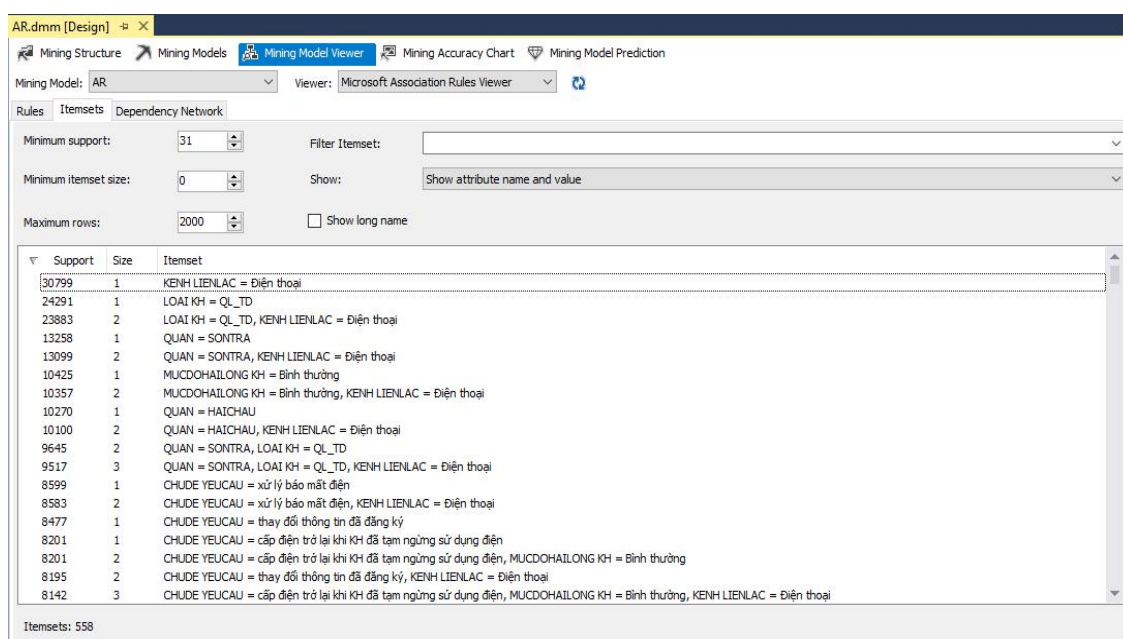
- **Luật 1:** Với mức độ hài lòng khách hàng là rất không hài lòng thì chủ đề yêu cầu là thay đổi hình thức thanh toán tiền điện với xác suất 94,3%.

- **Luật 2:** Với mức độ hài lòng khách hàng là bình thường thì chủ đề yêu cầu là cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện với xác suất 78,5%.

- **Luật 3:** Với mức độ hài lòng khách hàng là hài lòng thì chủ đề yêu cầu là thay đổi thông tin đã đăng ký với xác suất 66%.

## Itemsets

Itemsets cho biết các thông tin quan trọng của luật kết hợp như Support (độ hỗ trợ của luật kết hợp), Size (Số items trong Itemsets):



Support	Size	Itemset
30799	1	KENH LIENLAC = Điện thoại
24291	1	LOAI KH = QL_TD
23883	2	LOAI KH = QL_TD, KENH LIENLAC = Điện thoại
13258	1	QUAN = SONTRA
13099	2	QUAN = SONTRA, KENH LIENLAC = Điện thoại
10425	1	MUCDOHAILONG KH = Bình thường
10357	2	MUCDOHAILONG KH = Bình thường, KENH LIENLAC = Điện thoại
10270	1	QUAN = HAICHAU
10100	2	QUAN = HAICHAU, KENH LIENLAC = Điện thoại
9645	2	QUAN = SONTRA, LOAI KH = QL_TD
9517	3	QUAN = SONTRA, LOAI KH = QL_TD, KENH LIENLAC = Điện thoại
8599	1	CHUDE YEUCAU = xử lý báo mất điện
8583	2	CHUDE YEUCAU = xử lý báo mất điện, KENH LIENLAC = Điện thoại
8477	1	CHUDE YEUCAU = thay đổi thông tin đã đăng ký
8201	1	CHUDE YEUCAU = cấp điện trở lại khi KH đã tạm ngừng sử dụng điện
8201	2	CHUDE YEUCAU = cấp điện trở lại khi KH đã tạm ngừng sử dụng điện, MUCDOHAILONG KH = Bình thường
8195	2	CHUDE YEUCAU = thay đổi thông tin đã đăng ký, KENH LIENLAC = Điện thoại
8142	3	CHUDE YEUCAU = cấp điện trở lại khi KH đã tạm ngừng sử dụng điện, MUCDOHAILONG KH = Bình thường, KENH LIENLAC = Điện thoại

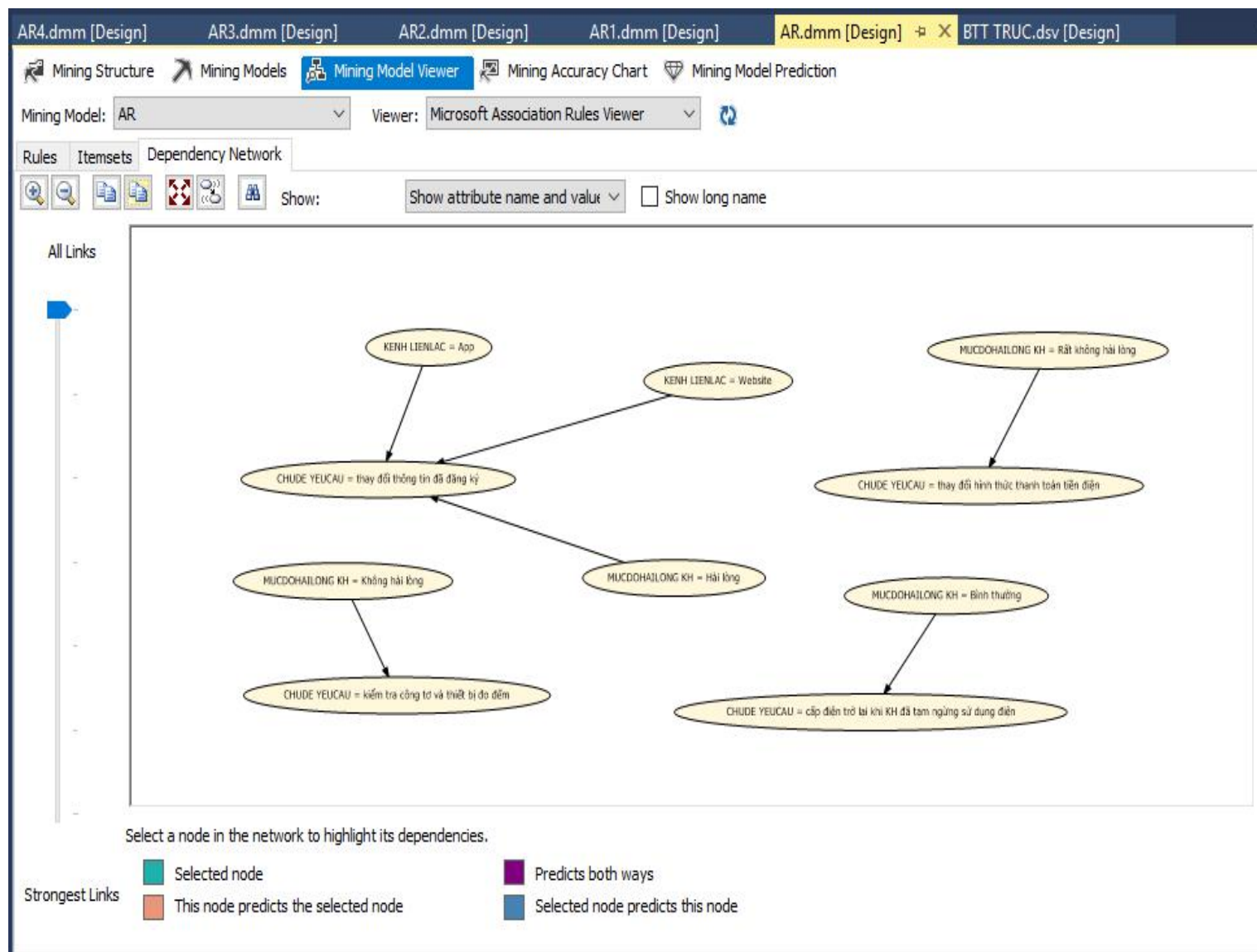
Itemsets: 558

**Hình 4.16. Itemsets của mô hình**

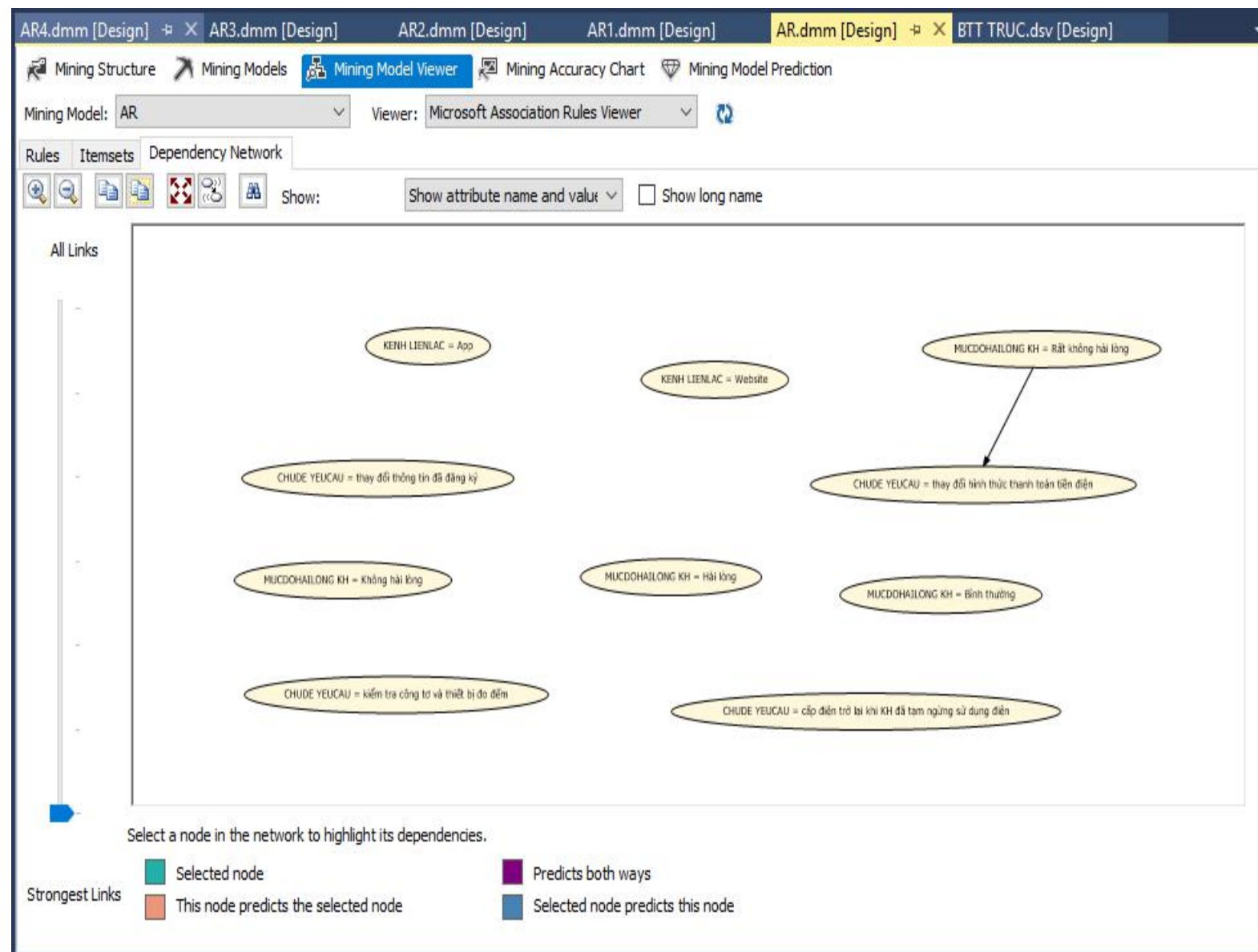
## Mạng phụ thuộc (Dependency Network)

- Sử dụng Dependency Network cho phép hiểu được sự tác động của các items khác nhau trong Model. Mỗi Node trong Dependency Net thể hiện một Item, bằng cách chọn một item, giao diện sẽ hiện lên các items khác được xác định bởi Item đã chọn (hoặc dùng để xác định Item đã chọn) trong model.

- Có thể kéo thanh trượt bên trái để xem các mức độ kết hợp (mạnh hay yếu) giữ các Items trong model.



**Hình 4.17a. Mạng phụ thuộc.**



**Hình 4.17b. Mạng phụ thuộc.**

## KẾT LUẬN

Bài báo cáo đã nghiên cứu và trình bày được những khái niệm lý thuyết cơ bản của công cụ khai phá dữ liệu Microsoft SQL Server 2014, công cụ trực quan hóa dữ liệu Power BI, đồng thời ứng dụng luật kết hợp để xây dựng mô hình khai phá dữ liệu. Thêm vào đó, việc nghiên cứu mức độ hài lòng của khách hàng là vô cùng cần thiết, nó giúp cho công ty có thể nhận ra sản phẩm dịch vụ của mình đang đứng ở đâu trên thị trường, chính sách chăm sóc khách hàng và chế độ hậu đãi đã tốt hay chưa.

Theo kết quả phân tích thì hầu hết khách hàng yêu cầu thuộc loại khách hàng Quản lý – tiêu dùng tại quận Sơn Trà và Hải Châu thông qua 2 kênh liên lạc phổ biến là Điện thoại và Website. Ngược lại, kênh liên lạc App, Chat và Email được ít người sử dụng hơn. Nhìn chung phần lớn khách hàng thường yêu cầu về các chủ đề như là: thay đổi thông tin đã đăng ký, xử lý báo mất điện, cấp điện trở lại khi khách hàng đã tạm ngừng sử dụng điện.

Về mức độ hài lòng sau khi được giải quyết các yêu cầu thì hầu hết khách hàng cảm thấy hài lòng về chủ đề thay đổi thông tin đã đăng ký. Tuy nhiên, về yêu cầu thay đổi hình thức thanh toán tiền điện thì phần lớn khách hàng cảm thấy không hài lòng.

Để nâng cao chất lượng phục vụ khách hàng, công ty Điện lực cần quan tâm và cải thiện những dịch vụ này tốt hơn để giảm thiểu phản hồi hay khó khăn của khách hàng khi sử dụng dịch vụ điện. Từ đó gia tăng thỏa mãn nhu cầu và đem lại sự thuận tiện cho khách hàng. Công ty có thể gia tăng chất lượng và nhân viên hỗ trợ chẳng hạn như tăng lượng nhân viên trực điện thoại, đầu tư trang thiết bị hiện đại cho tất cả các điểm dịch vụ nhằm nâng cao chất lượng dịch vụ cung cấp cho khách hàng, cố gắng giải quyết phản hồi một cách triệt để, hạn chế phản hồi vượt mức. Cần cải tiến các thủ tục giải quyết và tiếp nhận phản hồi của khách hàng, từ đó chất lượng giải quyết phản hồi được nâng cao hơn, rút ngắn thời gian trả lời phản hồi cho khách hàng. Thường xuyên đánh giá, thống kê tỷ lệ phản hồi của khách hàng để có giải pháp trả lời, hướng dẫn thỏa đáng. Thêm vào đó, cần quảng bá thêm các kênh liên lạc như App, Chat, Email để tránh tình trạng quá tải khi khách hàng sử dụng kênh liên lạc Điện thoại và Website quá nhiều. Đồng thời, thực hiện thêm các cuộc khảo sát độ hài lòng của khách hàng nhằm đảm bảo chất lượng về việc cải thiện các dịch vụ chăm sóc khách hàng.

## TÀI LIỆU THAM KHẢO

- [1] “Khai phá dữ liệu là gì ?,” 2017. [Trực tuyến]. Available: <https://it.die.vn/d/data-mining-khai-pha-du-lieu/>.
- [2] N. V. Chức, “Business Intelligence là gì ?,” 2010. [Trực tuyến]. Available: <http://bis.net.vn/forums/t/121.aspx>.
- [3] N. V. Chức, “Thuật toán Apriori khai phá luật kết hợp,” 2011. [Trực tuyến]. Available: <http://bis.net.vn/forums/t/389.aspx>.
- [4] “Tổng quan về SQL Server,” 2016. [Trực tuyến]. Available: <https://timoday.edu.vn/bai-1-tong-quan-ve-sql-server/>.
- [5] "SQL Server Analysis Services," 2011. [Online]. Available: <https://www.mssqltips.com/sqlservertutorial/2000/sql-server-analysis-services-ssas-tutorial/>.



**TRƯỜNG ĐẠI HỌC KINH TẾ**  
**KHOA THỐNG KÊ-TIN HỌC**

**CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM**  
**Độc lập - Tự do - Hạnh phúc**

**NHẬN XÉT CỦA GIÁNG VIÊN HƯỚNG DẪN**

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Đà Nẵng, Ngày ... tháng ... năm ...  
(Ký, ghi rõ họ và tên)

**NHẬN XÉT CỦA GIẢNG VIÊN PHẢN BIỆN**

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Đà Nẵng, Ngày ... tháng ... năm ...  
(Ký, ghi rõ họ và tên)