**BỘ MÔN HỆ THỐNG THÔNG TIN – KHOA CÔNG NGHỆ THÔNG TIN**

Logo

Description automatically generated**ĐẠI HỌC KHOA HỌC TỰ NHIÊN THÀNH PHỐ HỒ CHÍ MINH, ĐẠI HỌC QUỐC GIA TP HCM**

Sinh viên thực hiện: TTKD-28

GV phụ trách: Hồ Thị Hoàng Vy

  - HTTT PV TTKD

HỌC KỲ I – NĂM HỌC 2021-2022

**Báo cáo bài tập mining sử dụng association rule**

Danh sách thành viên

|  |  |
| --- | --- |
| Họ tên | MSSV |
| Võ Ngọc Đức | 18120327 |
| Phạm Ngọc Điệp | 18120316 |
| Nguyễn Văn Hải | 18120350 |
| Thái Vĩnh Toàn | 1712824 |

1. Xây dựng mô hình:

* Vấn đề: Dựa vào mining dữ liệu để thực hiện tìm ra các models thường được bán cùng nhau.
* Mô hình lựa chọn: Mô hình luật kết hợp.
* Chuẩn bị dữ liệu: Dữ liệu dùng để xây dựng mô hình phát hiện luật kết hợp là 2 view chứa dữ liệu về bán hàng (xe đạp) của kho dữ liệu AdventureWorksDW là vAssocSeqOrders – chứa thông tin về orders, customer và vAssocSeqLineItems- chứa nhiều dòng thông tin của các order này, các sản phẩm được mua.
* Cài đặt project:
  + B1: Tạo project Analysis services multidimentional and data mining project trên visual studio.
  + B2: Kết nối đến data source, tạo data source views chứa 2 view vAssocSeqOrders và vAssocSeqLineItems.

A screenshot of a computer

Description automatically generated

* + Với AdventureWorksDW 2019 thì giữa 2 view chưa có quan hệ. Nên trước khi thực hiện mining cần tạo mới 1 quan hệ giữa 2 view. Bằng cách chuột phải vAssocSeqLineItems. Chọn new relationship. Thực hiện tham chiếu qua OrderNumber. (cần thực hiện set logical primary key cho vAssocSeqLineItems trước)

Graphical user interface, text, application, email

Description automatically generated

* + B3: Tạo mining
  + Graphical user interface, text, application

    Description automatically generated
  + Lựa chọn dữ liệu sẽ mining dựa vào dữ liệu quan hệ hoặc warehold đã tạo sẵn.

Graphical user interface, text, application, email

Description automatically generated

* + Chọn mining model là microsoft association rules.

Graphical user interface, text, application, email

Description automatically generated

* + Chọn data source views

Graphical user interface, application

Description automatically generated

* + Lựa chọn case và nested. vAssocSeqOrders chứa dữ liệu bao hàm của dữ liệu liệu trong vAssocSeqLineItems (quan hệ 1 nhiều). Nên ta chọn vAssocSeqOrders là bảng case, vAssocSeqLineItems là bảng nested.

Graphical user interface, application

Description automatically generated

* + Xác định các thông tin nào sẽ là key, input, predict. Ở đây key với bảng case ta dùng ordernumber để match được key của nested là model. Với nhu cầu giải quyết vấn đề đặt ra là tìm các model sẽ được mua kèm với nhau thì input và predict sẽ là model.

Graphical user interface, table

Description automatically generated

* + Thiết lập content type.

Graphical user interface

Description automatically generated

* + Thiết lập tỷ lệ % dữ liệu dùng để test

Graphical user interface, text, application, email

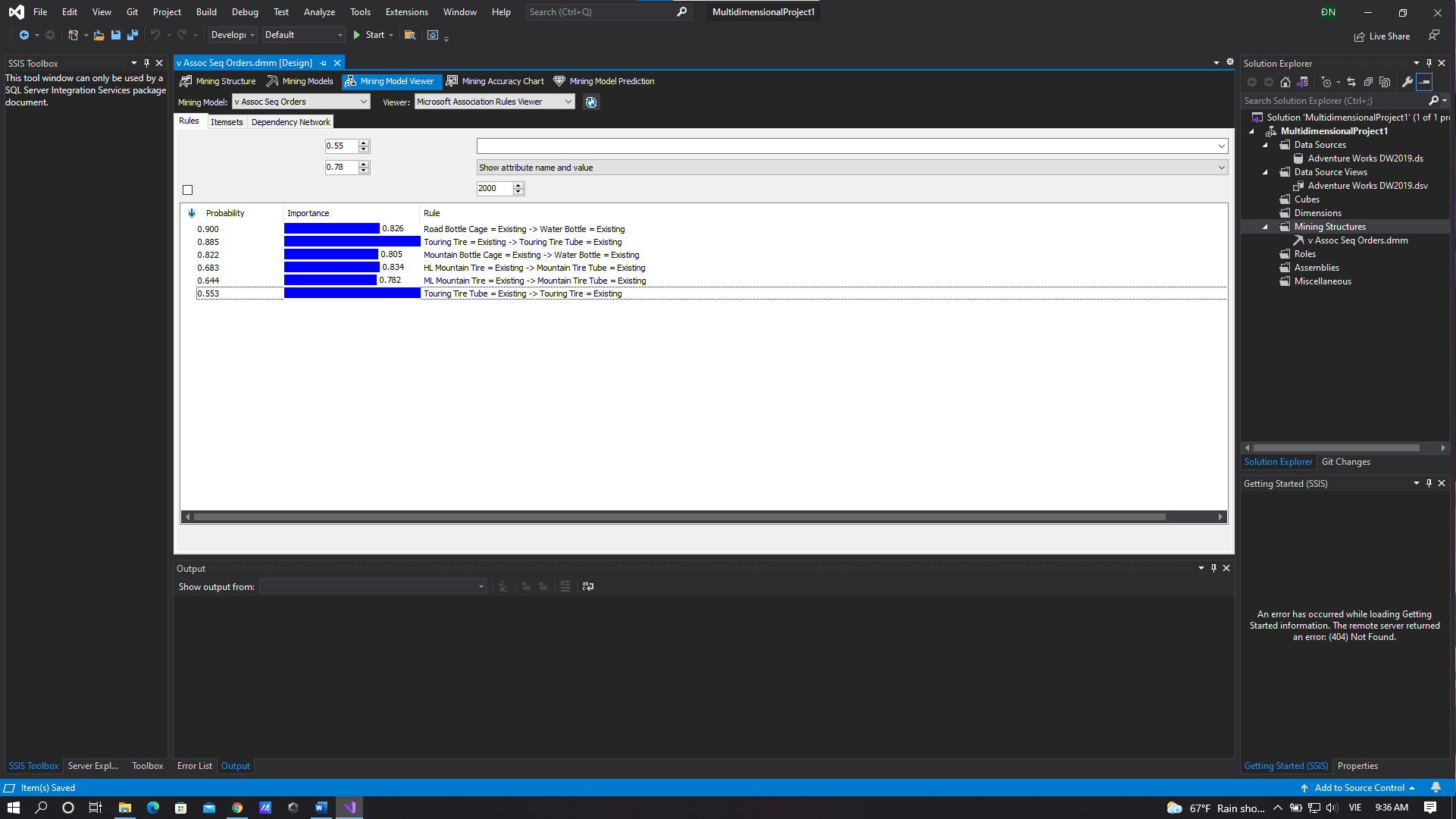
Description automatically generated

* + Cuối cùng đặt tên và nhấn finish.
  + Trước khi mining, cần thiết lập giá trị cho các tham số của thuật toán là Minimum-probability (xác suất thấp nhất có thể của rule) và minimum\_support (tỉ lệ thấp nhất mà tổng số trường hợp trong tập dữ liệu phải chứa model hoặc tập model này mới được đưa vào mô hình).

Graphical user interface

Description automatically generated

1. Đánh giá mô hình:
   * Các rule được rút ra:



A picture containing text, screenshot, monitor

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

* + Khi thay đổi các chỉ số support và probability thì sẽ thay đổi kết quả các luật nhận được sau quá trình mining bằng mô hình luật kết hợp. Các chỉ số này càng cao thì đem lại các luật càng có độ chính xác cao hơn, và giảm đi số lượng luật được sinh ra.
  + Với trường hợp trên thì ta cài đặt minimum\_support = 0.03 và minimum\_probability = 0.5.
  + Khi giảm các chỉ số này thì số luật được rút ra sẽ tăng lên và giảm đi độ chính xác của các luật.
  + Kết quả minh họa khi giảm minimum\_support = 0.01 và minimum\_probability = 0.3.

A picture containing chart

Description automatically generated

Graphical user interface, text, application

Description automatically generated

1. Predict:

* Case 1: Thực hiện dự đoán thử với hàm vừa được tạo, với input vào sẽ là road bottle cage thì 1 sản phẩm sẽ được mua là gì?

Graphical user interface, text, application

Description automatically generated

* Kết quả sẽ là mua kèm với water bottle.

Graphical user interface, application

Description automatically generated

* Case 2: Nếu khách hàng mua road bottle cage cùng touring-1000 thì 3 sản phẩm có thể được mua kèm là gì?

A screenshot of a computer

Description automatically generated with medium confidence

* Kết quả: Graphical user interface, application

  Description automatically generated