جامعة الأمير سلطان
PRINCE SULTAN UNIVERSITY

CCIS كلية علوم الحاسب والمعلومات
COLLEGE OF COMPUTER & INFORMATION SCIENCES

Assignment 3

Prepared by

Hussam Aldossary

Supervised by

Dr. Esam Mohammed Othman

Table of Contents

1. INTRODUCTION 3
   1.1. PURPOSE 3
   1.2. SCOPE 3

2. DATASET ANALYSIS AND PREPARATION 5
   2.1. Nursery Data 5
      2.1.1. Description 5
      2.1.2. features 5
      2.1.3. WEKA Analysis 7
      2.1.4. Results 17
   2.2. Online Retail Data 23
      2.2.1. Description 23
      2.2.2. features 23
      2.2.3. WEKA Analysis 24
      2.2.4. Results 32
3. REFERENCES 42

1 INTRODUCTION

1.1 Purpose

The purpose of this assignment is to apply practical data mining techniques—specifically Association Rule Mining—using the WEKA (Waikato Environment for Knowledge Analysis) toolset. Unlike classification, which predicts predefined labels, or clustering, which finds natural groupings, association mining explores relationships and patterns between items in a dataset, discovering rules of the form "if X then Y." Through hands-on analysis of two distinct datasets (Social Science and Business/Retail), this project aims to:

- Develop proficiency in unsupervised learning workflows for association tasks, from dataset acquisition to rule interpretation.
- Apply and understand the Apriori algorithm, a fundamental algorithm for finding frequent itemsets and generating association rules.
- Evaluate association performance using key metrics such as Support, Confidence, and Lift to determine the strength and interestingness of discovered rules.
- Demonstrate the utility of WEKA for exploratory data analysis and pattern discovery in both relational and transactional data contexts.

This report documents the methodology, analysis, and findings for each dataset, serving as a structured record of the analytical process and its outcomes.

1.2 SCOPE

The scope of this project is defined by the selection of two specific datasets and the use of the WEKA software for an Association task.

- **Datasets:**
    - **Nursery Data:** A multivariate dataset concerning the evaluation of applications for nursery schools in Ljubljana, Slovenia during the 1980s.
    - **Online Retail Data:** A transactional dataset containing all purchases occurring between 01/12/2010 and 09/12/2011 for a UK-based online retail store.
- **Tool:** All preprocessing and modeling tasks are performed within the WEKA Explorer environment.
- **Process:**
    - **Dataset Analysis:** Understanding the domain and attributes.
    - **Data Preparation:** Handling missing values, converting data to ARFF format, and for transactional data, aggregating items into a basket format suitable for association mining.
    - **Model Development:** Applying the Apriori algorithm to discover frequent itemsets and generate association rules.
    - **Evaluation:** Interpreting the generated rules using Support, Confidence, and Lift metrics. Using the "Classes to Associations" mode to evaluate how well rules predict a target class where applicable.

Deliverable: The final output is this comprehensive report, which includes dataset descriptions, methodological steps, analysis of results supported by WEKA screenshots, and a comparative discussion of findings.

Out of Scope:

- Deployment of models in a production environment.
- Extensive feature engineering beyond basic WEKA filters and necessary transactional reformatting.
- Comparative analysis with other data mining tools (e.g., R, Python's scikit-learn).
- In-depth theoretical derivation of the algorithms used.

2 DATASET ANALYSIS AND PREPARATION

This section provides a detailed overview of the two datasets selected from the UCI Machine Learning Repository for analysis. Each dataset represents an association machine learning task. For each dataset, a description of the problem domain, a summary of its features, and notes on data preparation for WEKA are provided.

2.1 Dataset 1: Association - Nursery Database

2.1.1 Description

The Nursery Database was derived from a hierarchical decision model originally developed to rank applications for nursery schools in Ljubljana, Slovenia, during the 1980s. The domain is Social Science. The goal of this analysis is to use association rule mining to discover combinations of family and child characteristics that lead to different application outcomes (recommended, priority, or not recommended).

The dataset was used during a period of excessive enrollment when rejected applications frequently needed an objective explanation. The original hierarchical model evaluated applications based on three subproblems: employment of parents and child's nursery, family structure and financial standing, and social and health picture of the family. For this analysis, the structural information has been removed, directly relating the final decision to eight input attributes. Successfully discovering strong association rules could reveal the implicit criteria used in the admission process and provide transparency for application decisions.

2.1.2 Features

The Nursery dataset contains **12,960 instances** and **9 attributes** (8 input attributes and 1 target class attribute). All attributes are categorical.

Feature Name	Data Type	Description	Possible Values
parents	Categorical	Parents' occupation level	usual, pretentious, great_pret
has_nurs	Categorical	Nursery arrangement of the child	proper, less_proper, improper, critical, very_crit
form	Categorical	Family structure	complete, completed, incomplete, foster

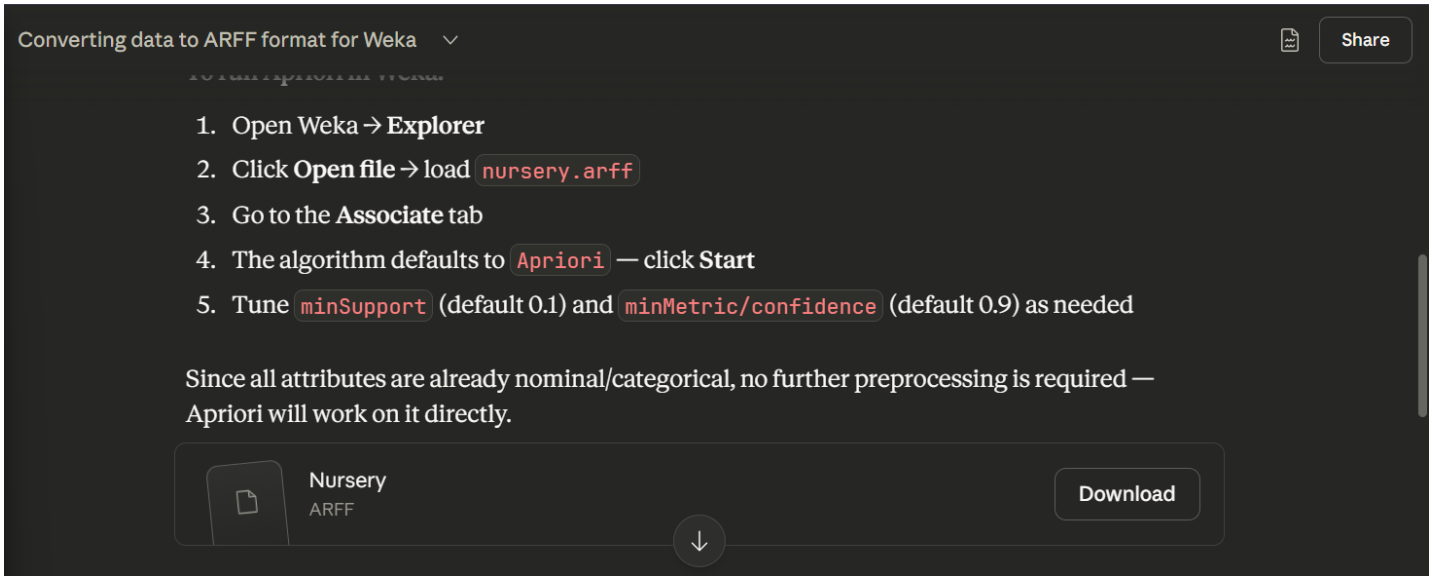
children	Categorical	Number of children	1, 2, 3, more
housing	Categorical	Housing conditions	convenient, less_conv, critical
finance	Categorical	Family financial status	convenient, inconv
social	Categorical	Social conditions	non-prob, slightly_prob, problematic
health	Categorical	Health conditions	recommended, priority, not_recom
class (Target)	**Categorical**	**Final evaluation result**	**recommended, priority, not_recom**

Key Dataset Facts:

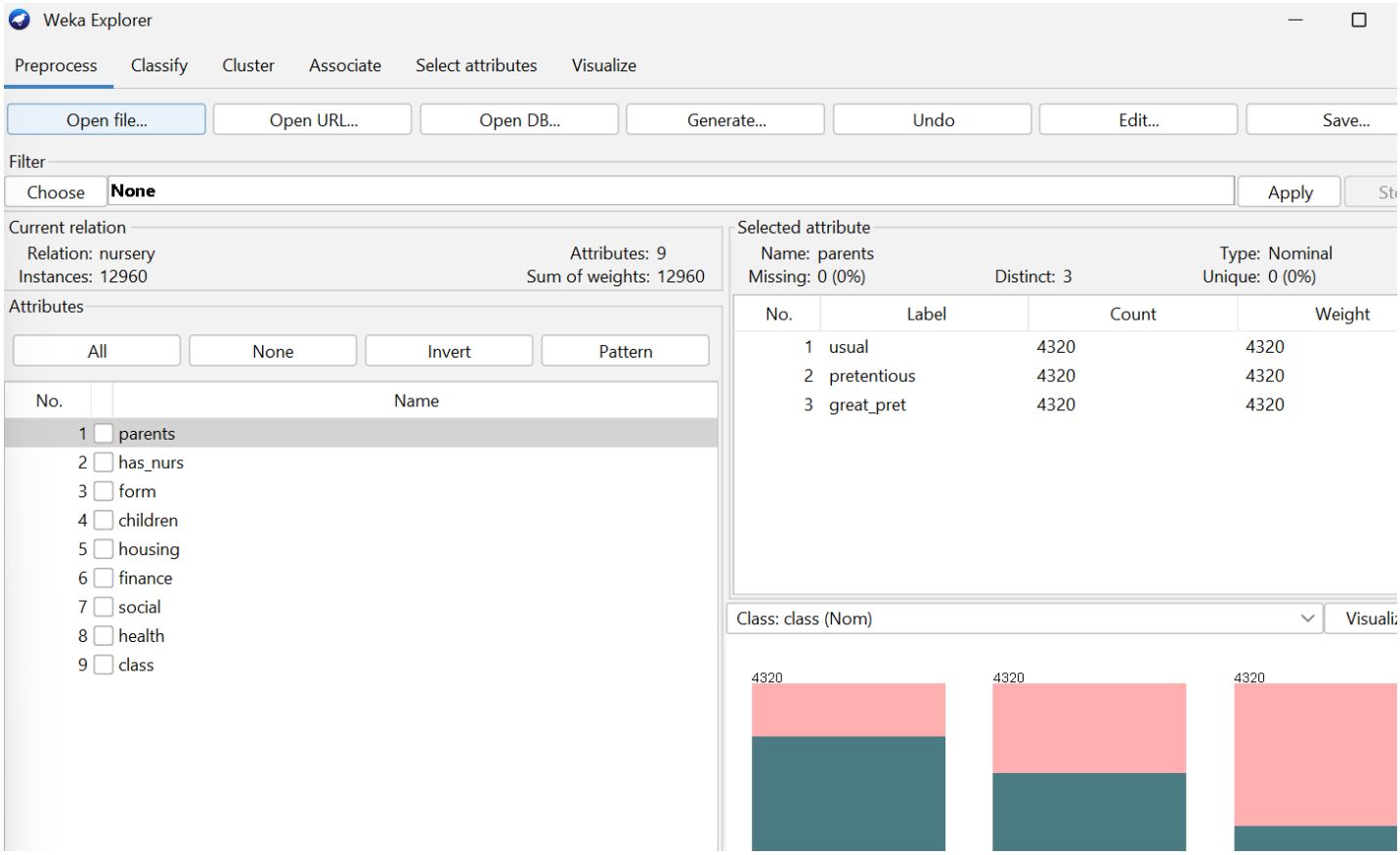
- Instances: 12,960
- Missing Values: None
- Features: Fully categorical/nominal
- Source File: nursery.data
- Domain: Social Science / Education

2.1.3 WEKA Analysis

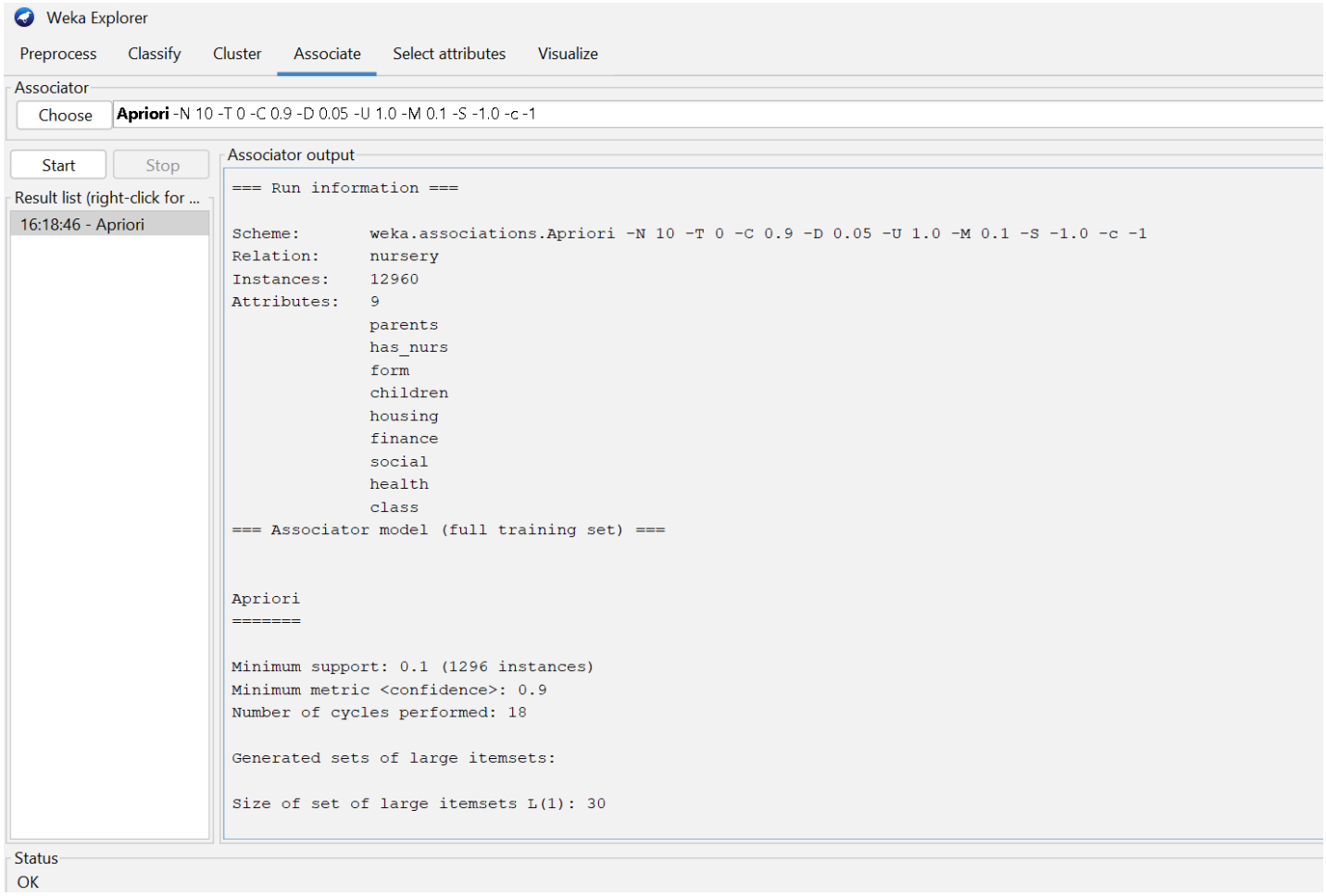
Using Claude to convert csv to arff

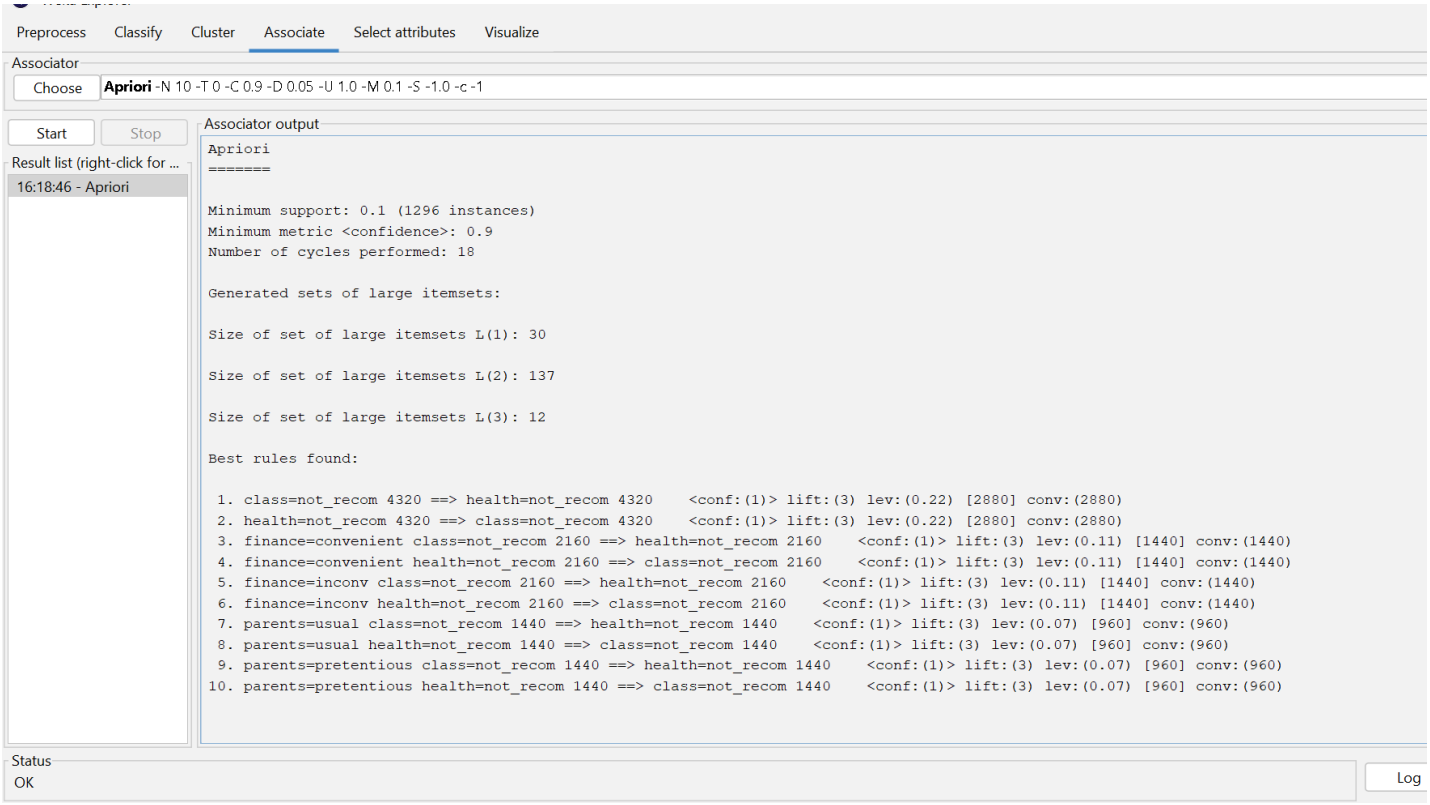


Loading data



Applying apriori (default parameters)





Now changing parameters to the following:
car = True

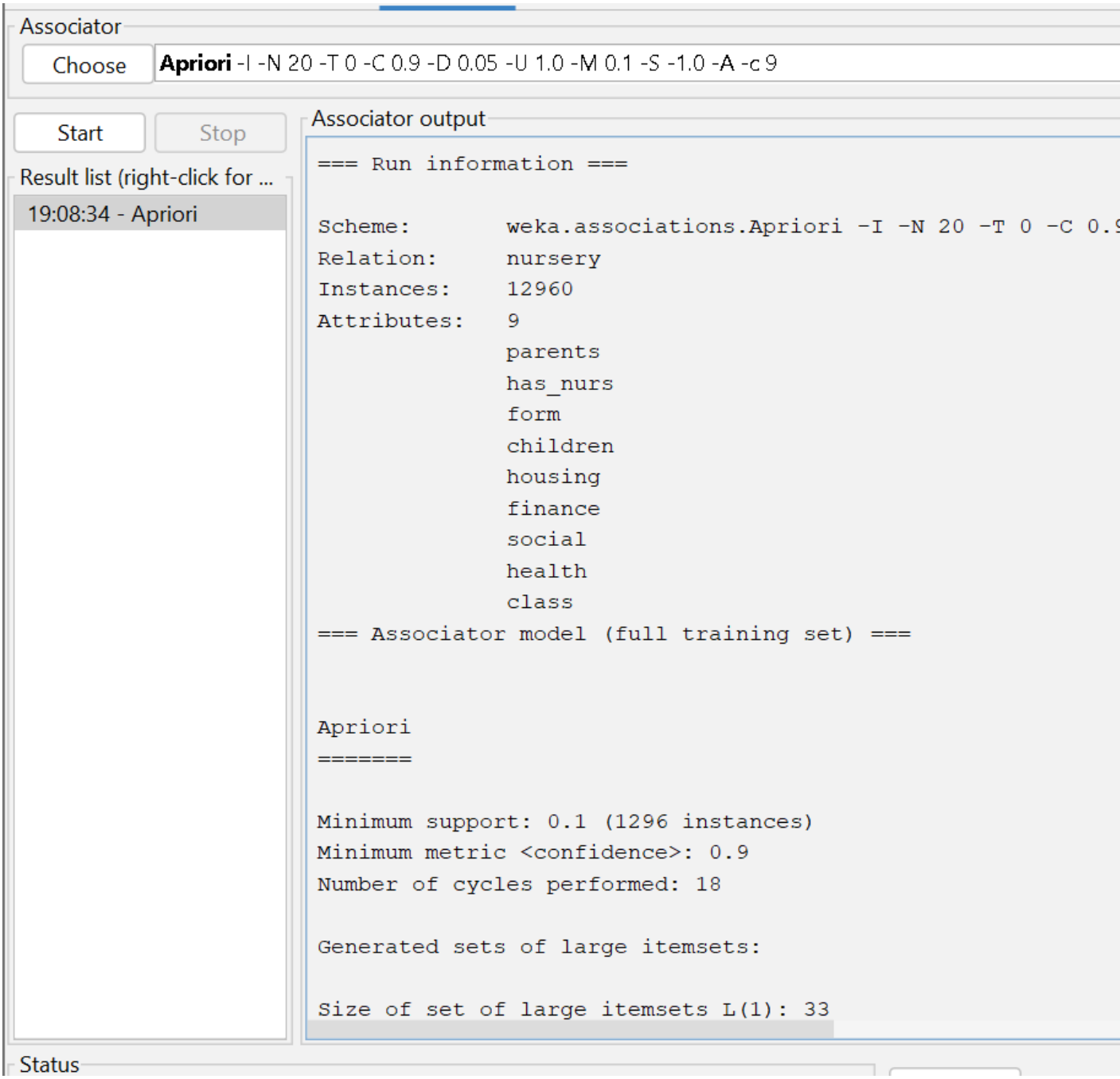
classIndex = 9

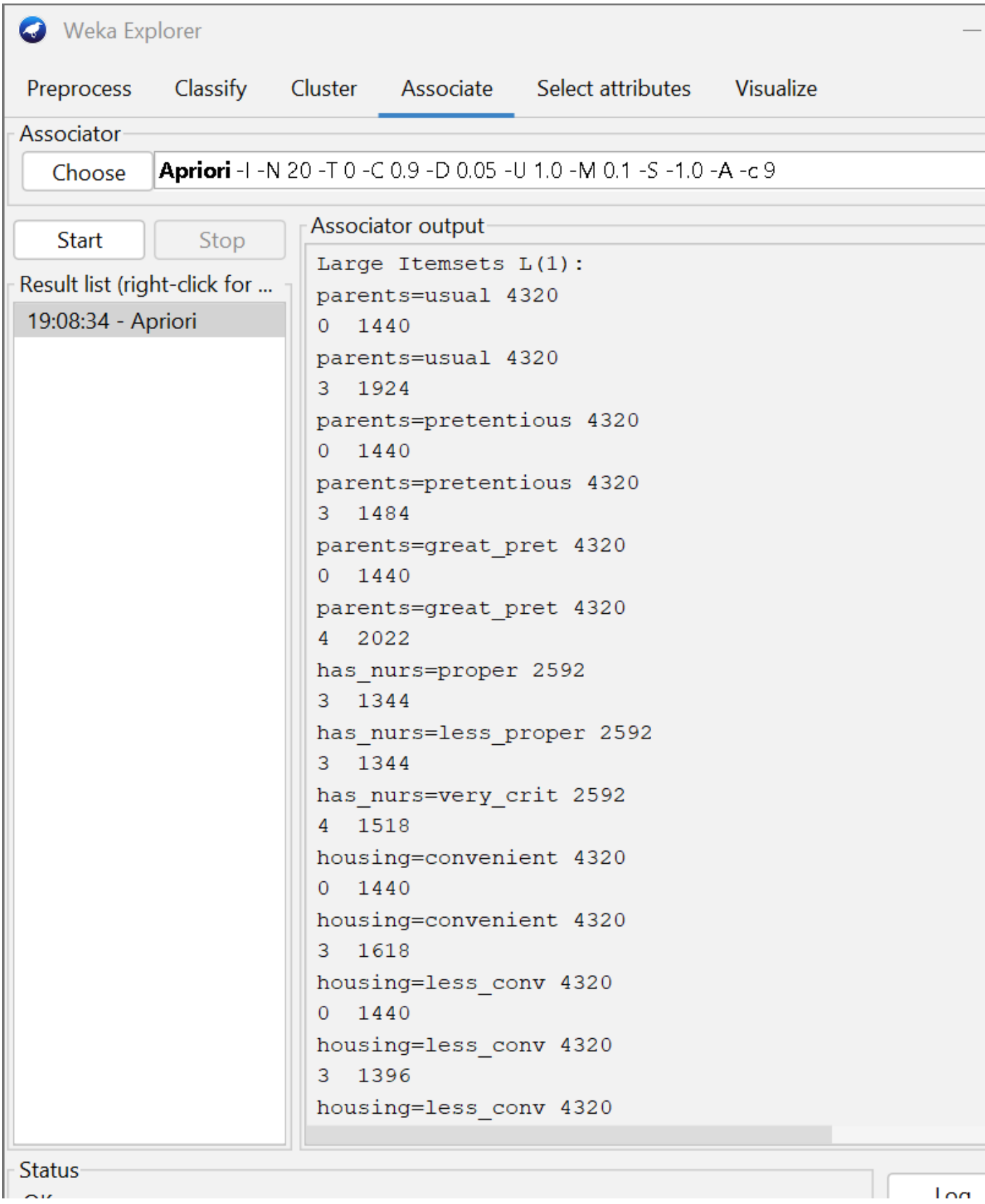
lowerBoundMinSupport = 0.1 (minimum support 10%)

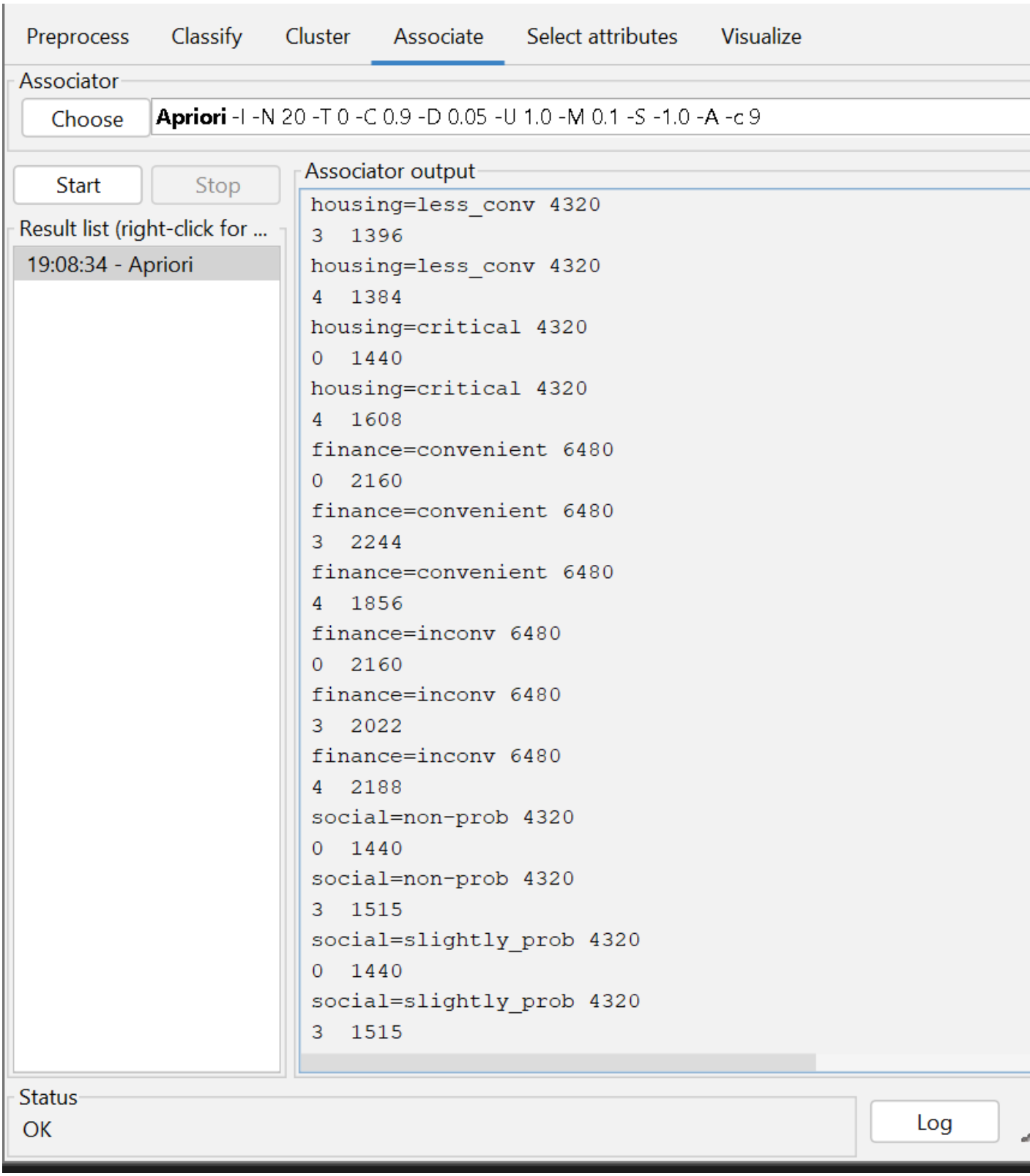
minMetric = 0.9 (minimum confidence 90%)

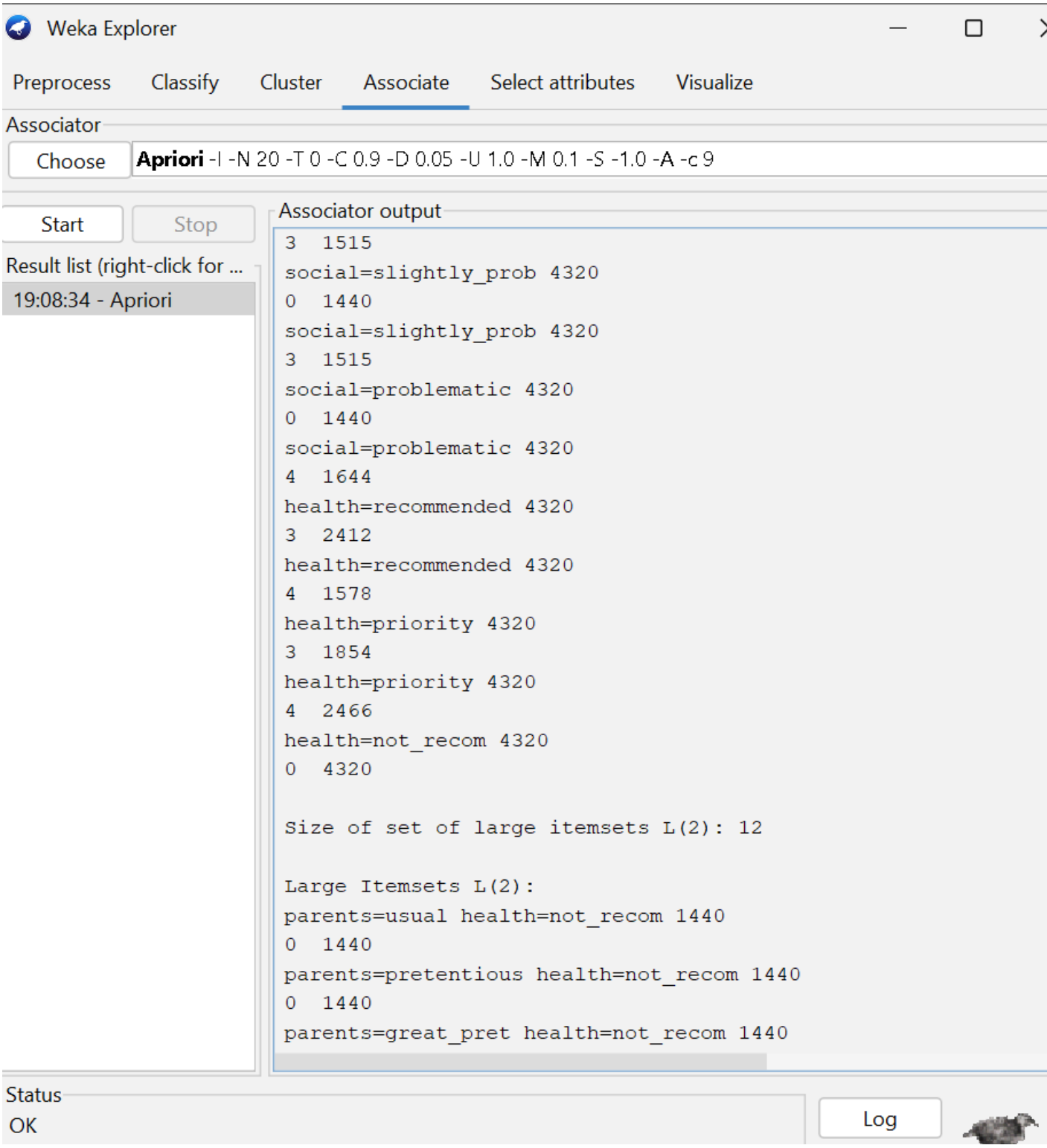
numRules = 20

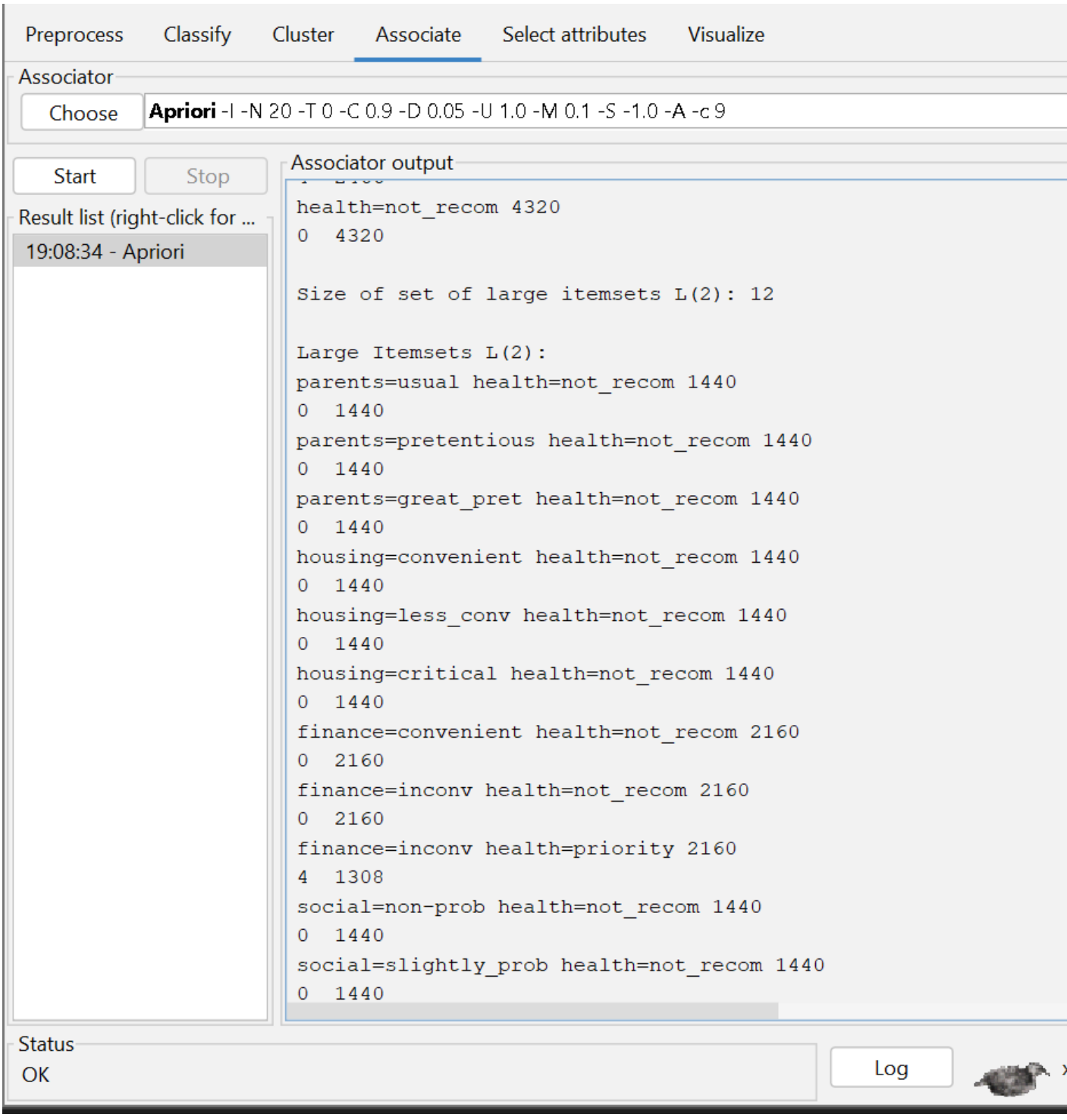
outputItemSets = True

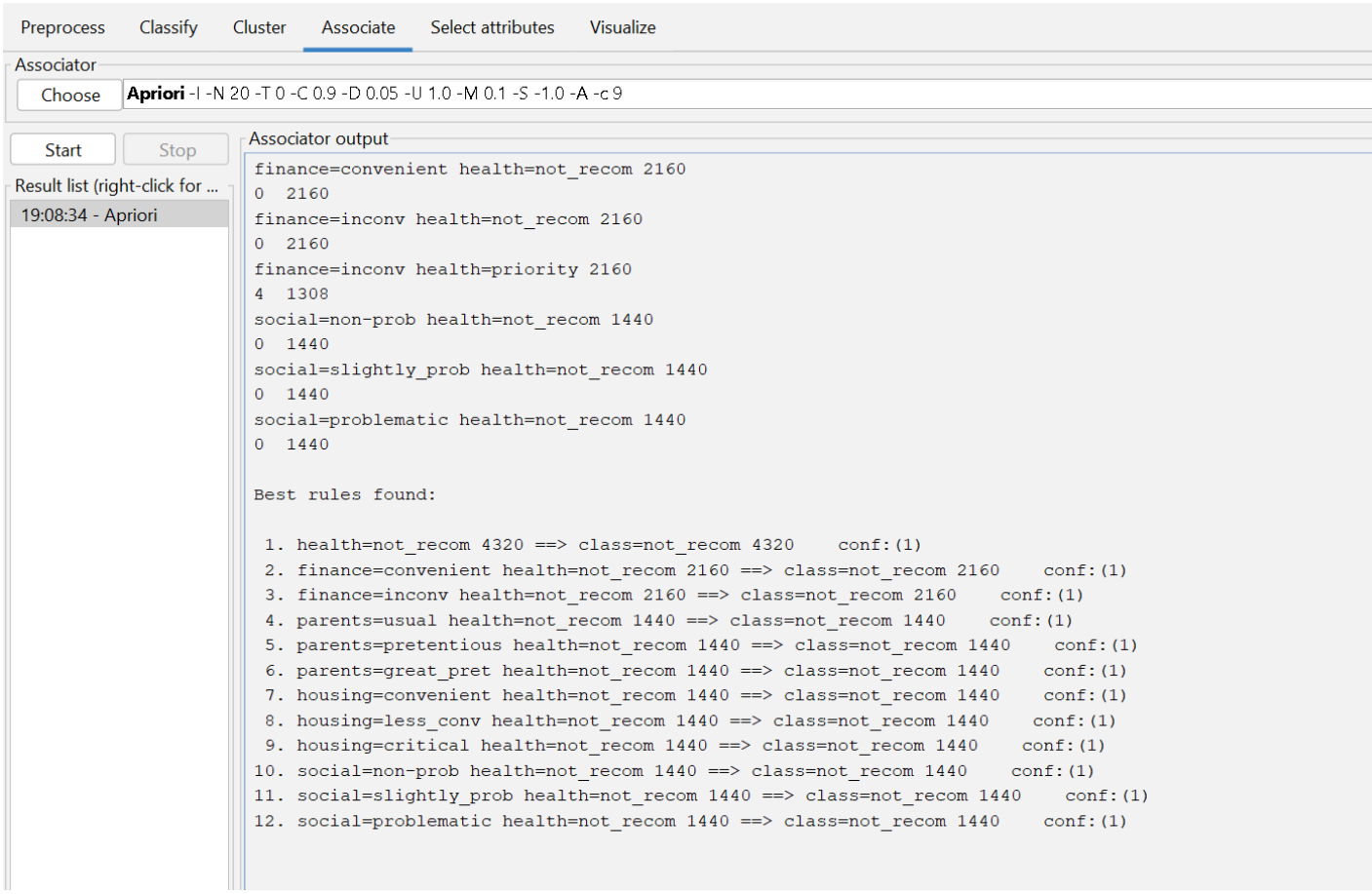



Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -I -N 20 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -A -c 9
Start
Stop
Result list (right-click for ...
19:08:34 - Apriori
Associator output
Large Itemsets L(1):
parents=usual 4320
0 1440
parents=usual 4320
3 1924
parents=pretentious 4320
0 1440
parents=pretentious 4320
3 1484
parents=great_pret 4320
0 1440
parents=great_pret 4320
4 2022
has_nurs=proper 2592
3 1344
has_nurs=less_proper 2592
3 1344
has_nurs=very_crit 2592
4 1518
housing=convenient 4320
0 1440
housing=convenient 4320
3 1618
housing=less_conv 4320
0 1440
housing=less_conv 4320
3 1396
housing=less_conv 4320
Status


Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -I -N 20 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -A -c 9
Start
Stop
Result list (right-click for ...
19:08:34 - Apriori
Associator output
housing=less_conv 4320
3 1396
housing=less_conv 4320
4 1384
housing=critical 4320
0 1440
housing=critical 4320
4 1608
finance=convenient 6480
0 2160
finance=convenient 6480
3 2244
finance=convenient 6480
4 1856
finance=inconv 6480
0 2160
finance=inconv 6480
3 2022
finance=inconv 6480
4 2188
social=non-prob 4320
0 1440
social=non-prob 4320
3 1515
social=slightly_prob 4320
0 1440
social=slightly_prob 4320
3 1515
Status
OK
Log


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -I -N 20 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -A -c 9
Start
Stop
Result list (right-click for ...
19:08:34 - Apriori
Associator output
3 1515
social=slightly_prob 4320
0 1440
social=slightly_prob 4320
3 1515
social=problematic 4320
0 1440
social=problematic 4320
4 1644
health=recommended 4320
3 2412
health=recommended 4320
4 1578
health=priority 4320
3 1854
health=priority 4320
4 2466
health=not_recom 4320
0 4320
Size of set of large itemsets L(2): 12
Large Itemsets L(2):
parents=usual health=not_recom 1440
0 1440
parents=pretentious health=not_recom 1440
0 1440
parents=great_pret health=not_recom 1440
Status
OK
Log


Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -I -N 20 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -A -c 9
Start
Stop
Result list (right-click for ...
19:08:34 - Apriori
Associator output
health=not_recom 4320
0 4320
Size of set of large itemsets L(2): 12
Large Itemsets L(2):
parents=usual health=not_recom 1440
0 1440
parents=pretentious health=not_recom 1440
0 1440
parents=great_pret health=not_recom 1440
0 1440
housing=convenient health=not_recom 1440
0 1440
housing=less_conv health=not_recom 1440
0 1440
housing=critical health=not_recom 1440
0 1440
finance=convenient health=not_recom 2160
0 2160
finance=inconv health=not_recom 2160
0 2160
finance=inconv health=priority 2160
4 1308
social=non-prob health=not_recom 1440
0 1440
social=slightly_prob health=not_recom 1440
0 1440
Status
OK
Log


Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -I -N 20 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -A -c 9
Start
Stop
Result list (right-click for ...
19:08:34 - Apriori
Associator output
finance=convenient health=not_recom 2160
0 2160
finance=inconv health=not_recom 2160
0 2160
finance=inconv health=priority 2160
4 1308
social=non-prob health=not_recom 1440
0 1440
social=slightly_prob health=not_recom 1440
0 1440
social=problematic health=not_recom 1440
0 1440
Best rules found:
1. health=not_recom 4320 ==> class=not_recom 4320 conf:(1)
2. finance=convenient health=not_recom 2160 ==> class=not_recom 2160 conf:(1)
3. finance=inconv health=not_recom 2160 ==> class=not_recom 2160 conf:(1)
4. parents=usual health=not_recom 1440 ==> class=not_recom 1440 conf:(1)
5. parents=pretentious health=not_recom 1440 ==> class=not_recom 1440 conf:(1)
6. parents=great_pret health=not_recom 1440 ==> class=not_recom 1440 conf:(1)
7. housing=convenient health=not_recom 1440 ==> class=not_recom 1440 conf:(1)
8. housing=less_conv health=not_recom 1440 ==> class=not_recom 1440 conf:(1)
9. housing=critical health=not_recom 1440 ==> class=not_recom 1440 conf:(1)
10. social=non-prob health=not_recom 1440 ==> class=not_recom 1440 conf:(1)
11. social=slightly_prob health=not_recom 1440 ==> class=not_recom 1440 conf:(1)
12. social=problematic health=not_recom 1440 ==> class=not_recom 1440 conf:(1)

2.1.4 results

Apriori Configuration Summary

Two experiments were conducted: one with the default parameters (no class index specified) and one with the class index set to attribute 9 (class) to generate class-association rules. The key settings and outcomes are summarised in Table 1.

Table 1: Apriori Parameter Settings and Performance Summary (Nursery Dataset)

Parameter	Default Run	Configured Run (Class-Association)
Class index	Not set	9 (class)
Minimum support	0.1 (1296 instances)	0.1 (1296 instances)
Minimum confidence	0.9	0.9
Number of cycles	18	18
Large itemsets L(1)	30	33
Large itemsets L(2)	137	12
Large itemsets L(3)	12	0

Number of rules generated	**10 (default limit)**	**12**
Rules predicting class=not_recom	**6**	**12**
Rules predicting class=priority / recommended	**0**	**0**

Discovered Association Rules

The configured run (class index = 9) produced 12 rules, all with perfect confidence (1.0) and high lift. Table 2 lists these rules together with their support and lift values.

Table 2: Top Association Rules for Nursery Data (Configured Run, Class Index = 9)

Rule	Antecedent	Consequent	Support (instances)	Support (%)	Confidence	Lift	Interpretation
1	health=not_recom	class=not_recom	4320	33.3 %	1.0	3.0	If health is “not recommended”, the application is always rejected.

2	**finance=convenient AND health=not_recom**	**class=not_recom**	**2160**	**16.7 %**	**1.0**	**3.0**	**Convenient finances together with poor health always lead to rejection.**
3	**finance=inconv AND health=not_recom**	**class=not_recom**	**2160**	**16.7 %**	**1.0**	**3.0**	**Inconvenient finances and poor health also guarantee rejection.**
4	**parents=usual AND health=not_recom**	**class=not_recom**	**1440**	**11.1 %**	**1.0**	**3.0**	**Usual parent occupation plus poor health ⇒ rejection.**
5	**parents=pretentious AND health=not_recom**	**class=not_recom**	**1440**	**11.1 %**	**1.0**	**3.0**	**Pretentious parents with poor health ⇒ rejection.**
6	**parents=great_pret AND health=not_recom**	**class=not_recom**	**1440**	**11.1 %**	**1.0**	**3.0**	**Great pretension parents with poor health ⇒ rejection.**

7	**housing=convenient AND health=not_recom**	**class=not_recom**	**1440**	**11.1 %**	**1.0**	**3.0**	**Convenient housing with poor health ⇒ rejection.**
8	**housing=less_conv AND health=not_recom**	**class=not_recom**	**1440**	**11.1 %**	**1.0**	**3.0**	**Less convenient housing with poor health ⇒ rejection.**
9	**housing=critical AND health=not_recom**	**class=not_recom**	**1440**	**11.1 %**	**1.0**	**3.0**	**Critical housing with poor health ⇒ rejection.**
10	**social=non-prob AND health=not_recom**	**class=not_recom**	**1440**	**11.1 %**	**1.0**	**3.0**	**Non-problematic social conditions with poor health ⇒ rejection.**
11	**social=slightly_prob AND health=not_recom**	**class=not_recom**	**1440**	**11.1 %**	**1.0**	**3.0**	**Slightly problematic social conditions with poor health ⇒ rejection.**

12	**social=problematic AND health=not_recom**	**class=not_recom**	**1440**	**11.1%**	**1.0**	**3.0**	**Problematic social conditions with poor health ⇒ rejection.**

Note: Lift is calculated as* confidence / P(consequent)*. Since* P(class=not_recom) = 4320/12960 = 1/3*, a confidence of 1.0 gives a lift of 3.0.

Frequent Itemsets (Configured Run)

The algorithm also generated frequent itemsets of size 1 and 2. Table 3 lists the most significant large itemsets from L(2) that involve the class attribute.

Table 3: Selected Large Itemsets of Size 2 (Configured Run)

Itemset	Support (instances)	Support (%)
parents=usual AND health=not_recom	**1440**	**11.1%**
parents=pretentious AND health=not_recom	**1440**	**11.1%**
parents=great_pret AND health=not_recom	**1440**	**11.1%**
housing=convenient AND health=not_recom	**1440**	**11.1%**
housing=less_conv AND health=not_recom	**1440**	**11.1%**
housing=critical AND health=not_recom	**1440**	**11.1%**

finance=convenient AND health=not_recom	**2160**	**16.7%**
finance=inconv AND health=not_recom	**2160**	**16.7%**
social=non-prob AND health=not_recom	**1440**	**11.1%**
social=slightly_prob AND health=not_recom	**1440**	**11.1%**
social=problematic AND health=not_recom	**1440**	**11.1%**

Observations

- **The default run (without a class index) produced symmetric rules, such as class=not_recom ⇒ health=not_recom and health=not_recom ⇒ class=not_recom, both with confidence 1.0 and lift 3.0. This confirms a perfect bidirectional relationship between the class label "not recommended" and the health attribute value "not recommended". Other rules from the default run also linked various attribute values (finance, parents, etc.) with either class or health, but the configured run makes the predictive patterns clearer.**
- **In the configured run, all 12 discovered rules predict class=not_recom and all include health=not_recom in the antecedent. This indicates that in this dataset, poor health is the single most powerful predictor of application rejection. No rules predicting the other two classes (priority or recommended) were found with the chosen minimum support (10%) and confidence (90%).**
- **The rules with finance=convenient or finance=inconv combined with health=not_recom have higher support (2160 instances) than those involving parents, housing, or social attributes (1440 instances). This suggests that financial standing, when paired with poor health, appears more frequently among rejected applications.**
- **All rules achieve a lift of 3.0, meaning they are three times more likely to occur than random chance—an exceptionally strong association.**
- **The absence of rules for priority and recommended classes implies that those outcomes are not associated with any single attribute or pair of attributes at the 10% support level; they may require more specific combinations (higher-order itemsets) or lower support thresholds to be discovered.**

2.2 Dataset 2: Association - Online Retail

2.2.1 Description

The Online Retail dataset contains all transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail company. The domain is Business and E-commerce. The goal of this analysis is to use association rule mining (market basket analysis) to discover relationships between products that are frequently purchased together in the same transaction.

This type of analysis is fundamental for understanding customer purchasing behavior, optimizing product placement, designing cross-selling strategies, and informing promotional decisions. The dataset includes approximately one year of transactions, capturing seasonal patterns and a wide variety of customer purchasing behaviors across different countries.

2.2.2 Features

The dataset contains 541,909 instances (individual line items) and 8 attributes. Features include a mix of categorical, integer, real, and date types.

Feature Name	Role	Data Type	Description	Missing Values
InvoiceNo	ID	Categorical	Unique invoice number for each transaction. Codes beginning with 'c' indicate cancellations.	no
StockCode	ID	Categorical	Unique product identifier code.	no
Description	Feature	Categorical	Name of the product.	no
Quantity	Feature	Integer	Number of units purchased per product in a transaction.	no

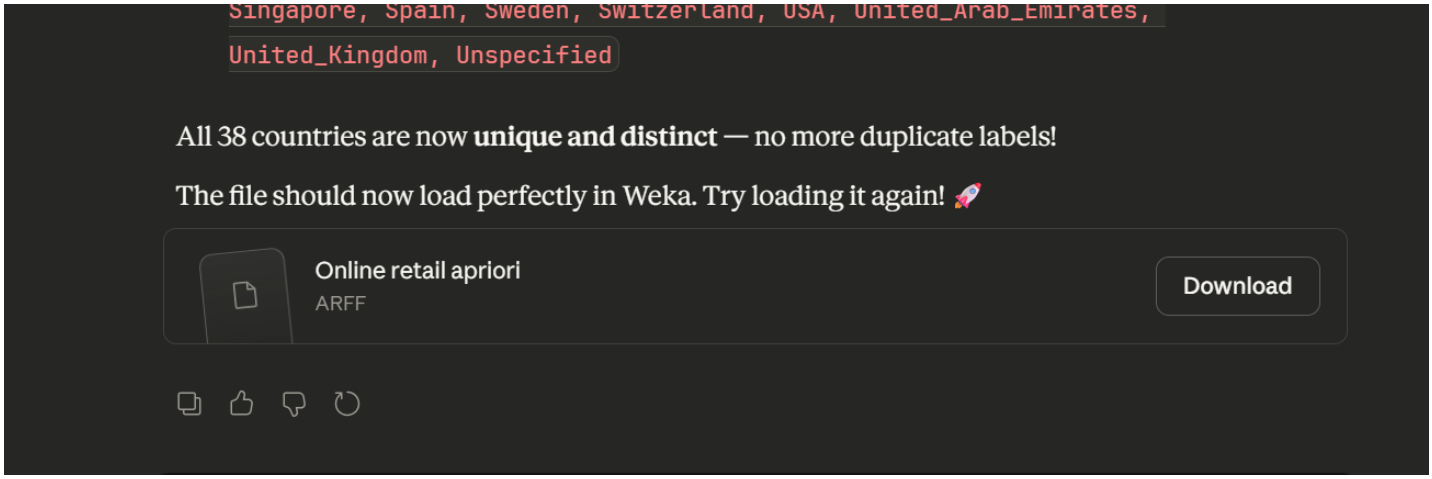
InvoiceDate	Feature	Date	Date and time when the transaction occurred.	no
UnitPrice	Feature	Continuous	Price per unit in sterling currency.	no
CustomerID	Feature	Categorical	Unique identifier for each customer.	no
Country	Feature	Categorical	Country where the customer is located.	no

Key Dataset Facts:

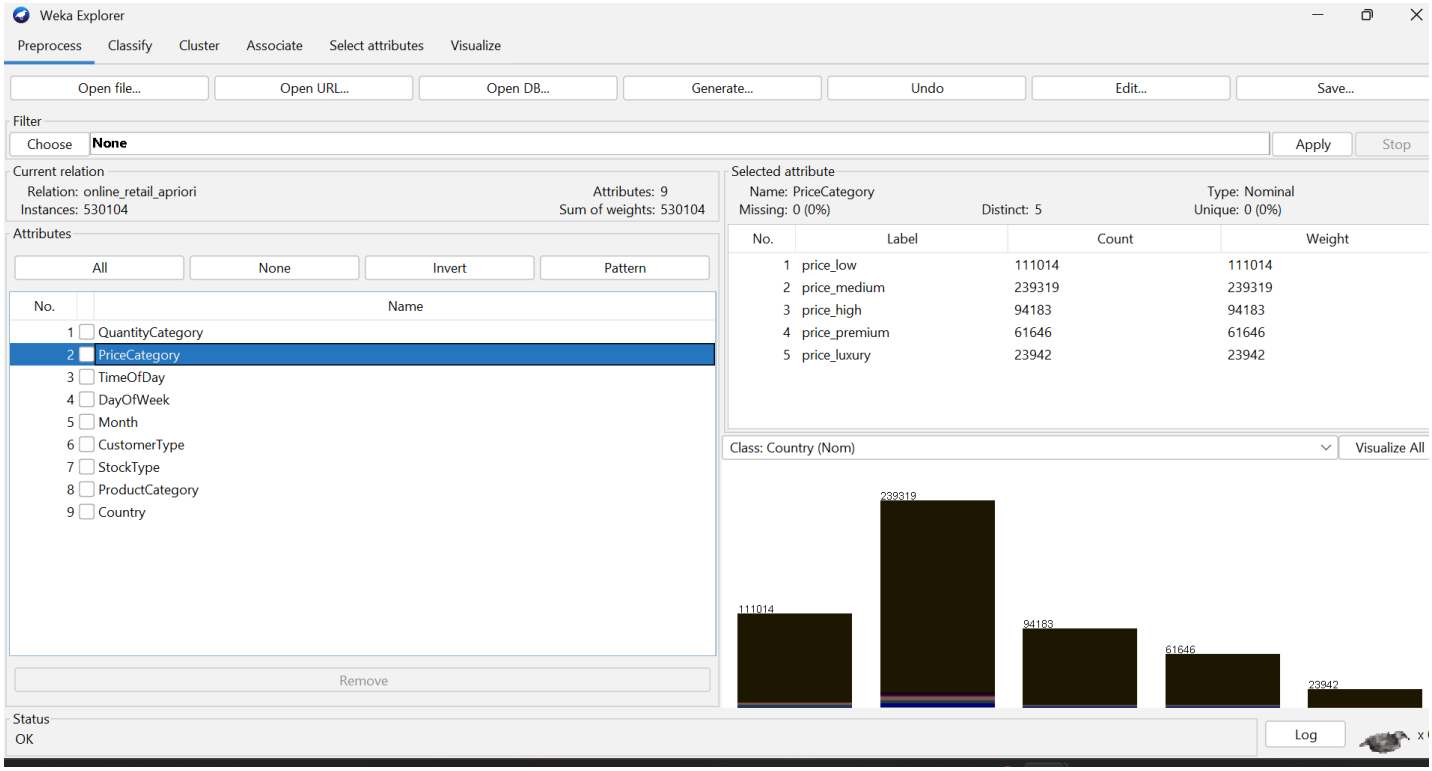
- Instances: 541,909 (line items)
- Missing Values: None reported in documentation
- Features: Mixed types (categorical, integer, real, date)
- Source File: Online Retail.xlsx
- Domain: Business / E-commerce

2.2.3 WEKA Analysis

Converting to arff using claude

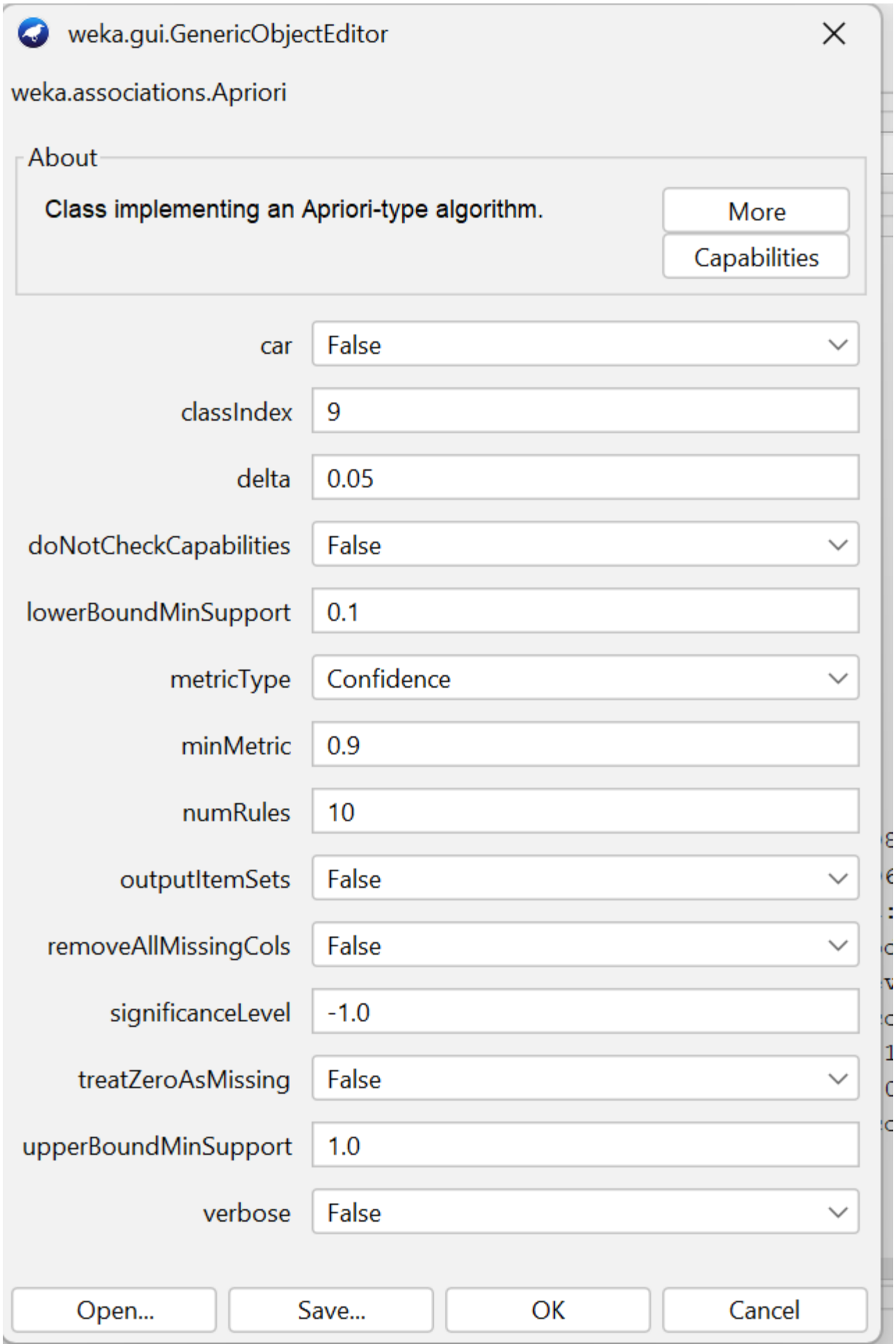


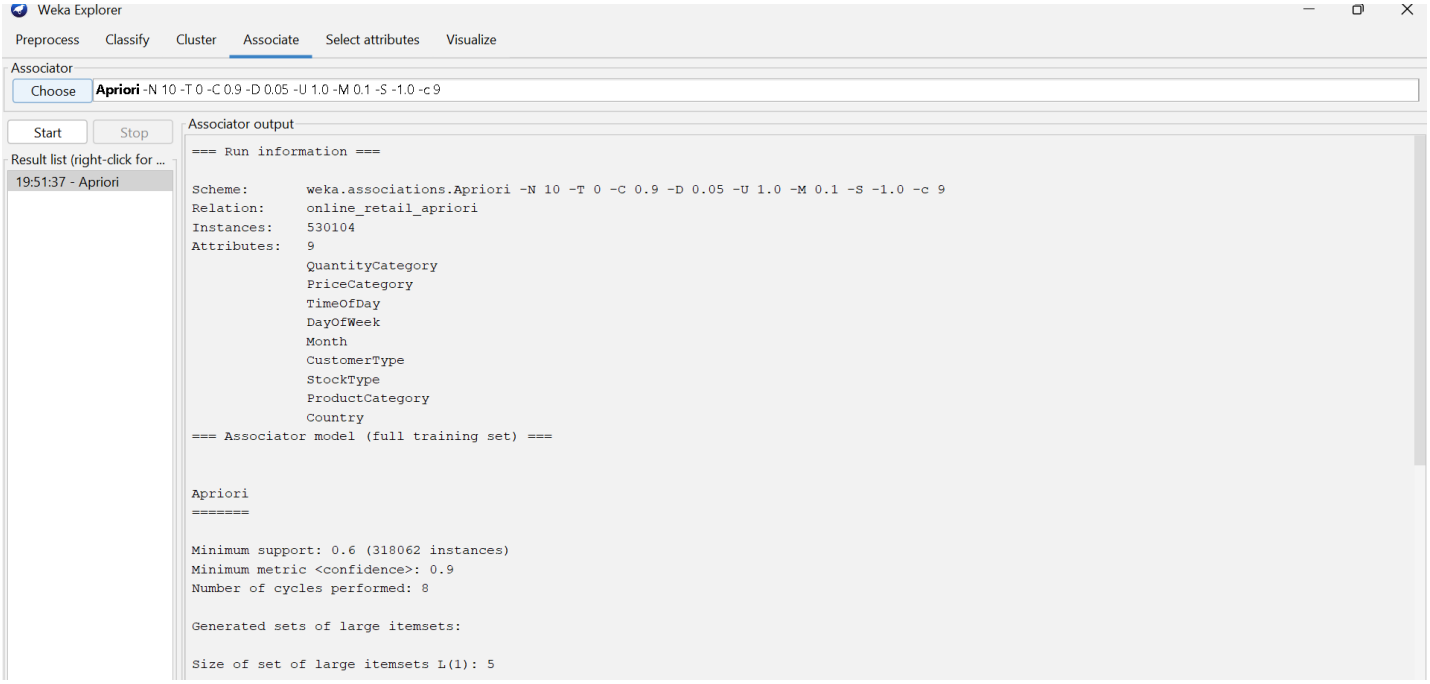
Loading data

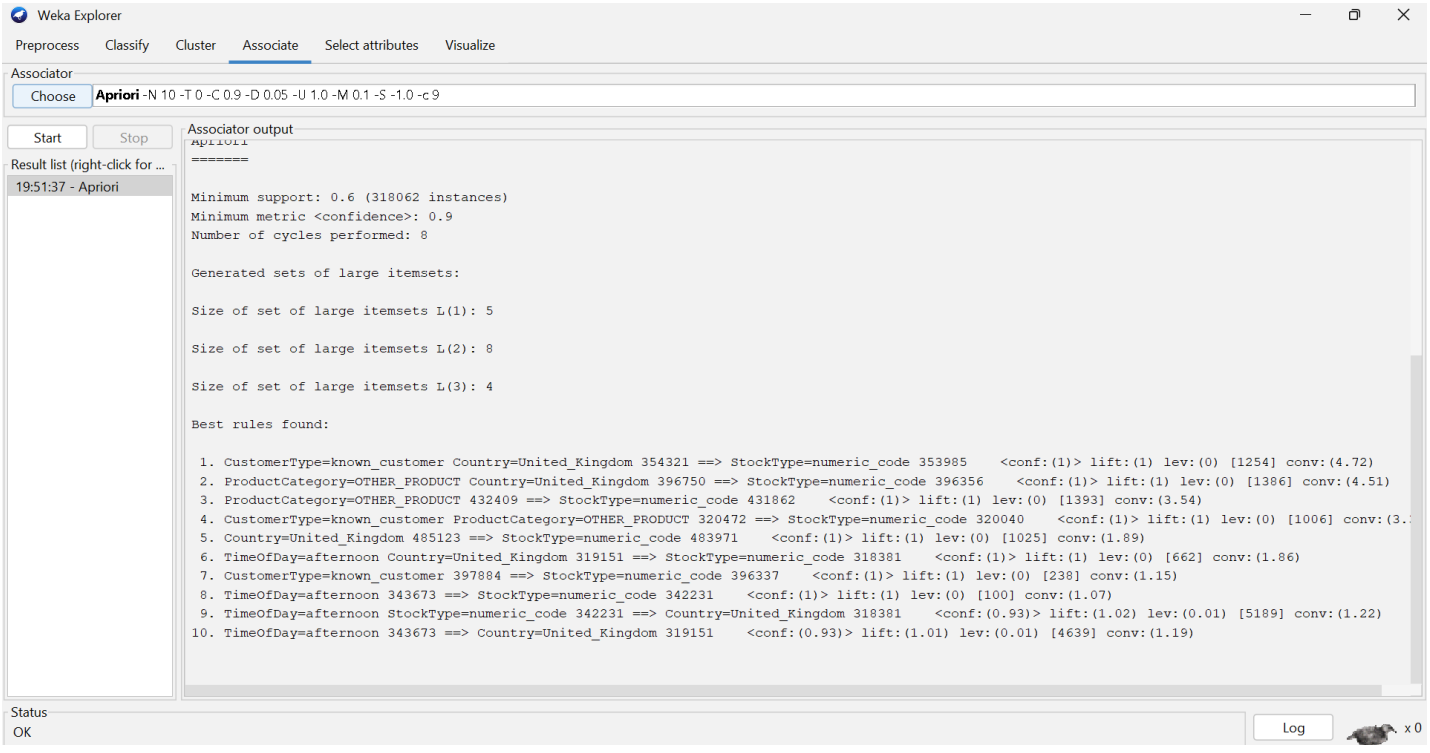


Preprocessing was done by Claude previously

Arpriori (default parameters)


weka.gui.GenericObjectEditor
weka.associations.Apriori
About
Class implementing an Apriori-type algorithm.
More
Capabilities
car
False
classIndex
9
delta
0.05
doNotCheckCapabilities
False
lowerBoundMinSupport
0.1
metricType
Confidence
minMetric
0.9
numRules
10
outputItemSets
False
removeAllMissingCols
False
significanceLevel
-1.0
treatZeroAsMissing
False
upperBoundMinSupport
1.0
verbose
False
Open...
Save...
OK
Cancel


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c 9
Start
Stop
Result list (right-click for ...
19:51:37 - Apriori
Associator output
=== Run information ===
Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c 9
Relation: online_retail_apriori
Instances: 530104
Attributes: 9
QuantityCategory
PriceCategory
TimeOfDay
DayOfWeek
Month
CustomerType
StockType
ProductCategory
Country
=== Associator model (full training set) ===
Apriori
=======
Minimum support: 0.6 (318062 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 8
Generated sets of large itemsets:
Size of set of large itemsets L(1): 5


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c 9
Start
Stop
Result list (right-click for ...
19:51:37 - Apriori
Associator output
=======
Minimum support: 0.6 (318062 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 8
Generated sets of large itemsets:
Size of set of large itemsets L(1): 5
Size of set of large itemsets L(2): 8
Size of set of large itemsets L(3): 4
Best rules found:
1. CustomerType=known_customer Country=United_Kingdom 354321 ==> StockType=numeric_code 353985 <conf:(1)> lift:(1) lev:(0) [1254] conv:(4.72)
2. ProductCategory=OTHER_PRODUCT Country=United_Kingdom 396750 ==> StockType=numeric_code 396356 <conf:(1)> lift:(1) lev:(0) [1386] conv:(4.51)
3. ProductCategory=OTHER_PRODUCT 432409 ==> StockType=numeric_code 431862 <conf:(1)> lift:(1) lev:(0) [1393] conv:(3.54)
4. CustomerType=known_customer ProductCategory=OTHER_PRODUCT 320472 ==> StockType=numeric_code 320040 <conf:(1)> lift:(1) lev:(0) [1006] conv:(3.
5. Country=United_Kingdom 485123 ==> StockType=numeric_code 483971 <conf:(1)> lift:(1) lev:(0) [1025] conv:(1.89)
6. TimeOfDay=afternoon Country=United_Kingdom 319151 ==> StockType=numeric_code 318381 <conf:(1)> lift:(1) lev:(0) [662] conv:(1.86)
7. CustomerType=known_customer 397884 ==> StockType=numeric_code 396337 <conf:(1)> lift:(1) lev:(0) [238] conv:(1.15)
8. TimeOfDay=afternoon 343673 ==> StockType=numeric_code 342231 <conf:(1)> lift:(1) lev:(0) [100] conv:(1.07)
9. TimeOfDay=afternoon StockType=numeric_code 342231 ==> Country=United_Kingdom 318381 <conf:(0.93)> lift:(1.02) lev:(0.01) [5189] conv:(1.22)
10. TimeOfDay=afternoon 343673 ==> Country=United_Kingdom 319151 <conf:(0.93)> lift:(1.01) lev:(0.01) [4639] conv:(1.19)
Status
OK
Log
x 0

Arpriori (different parameters)

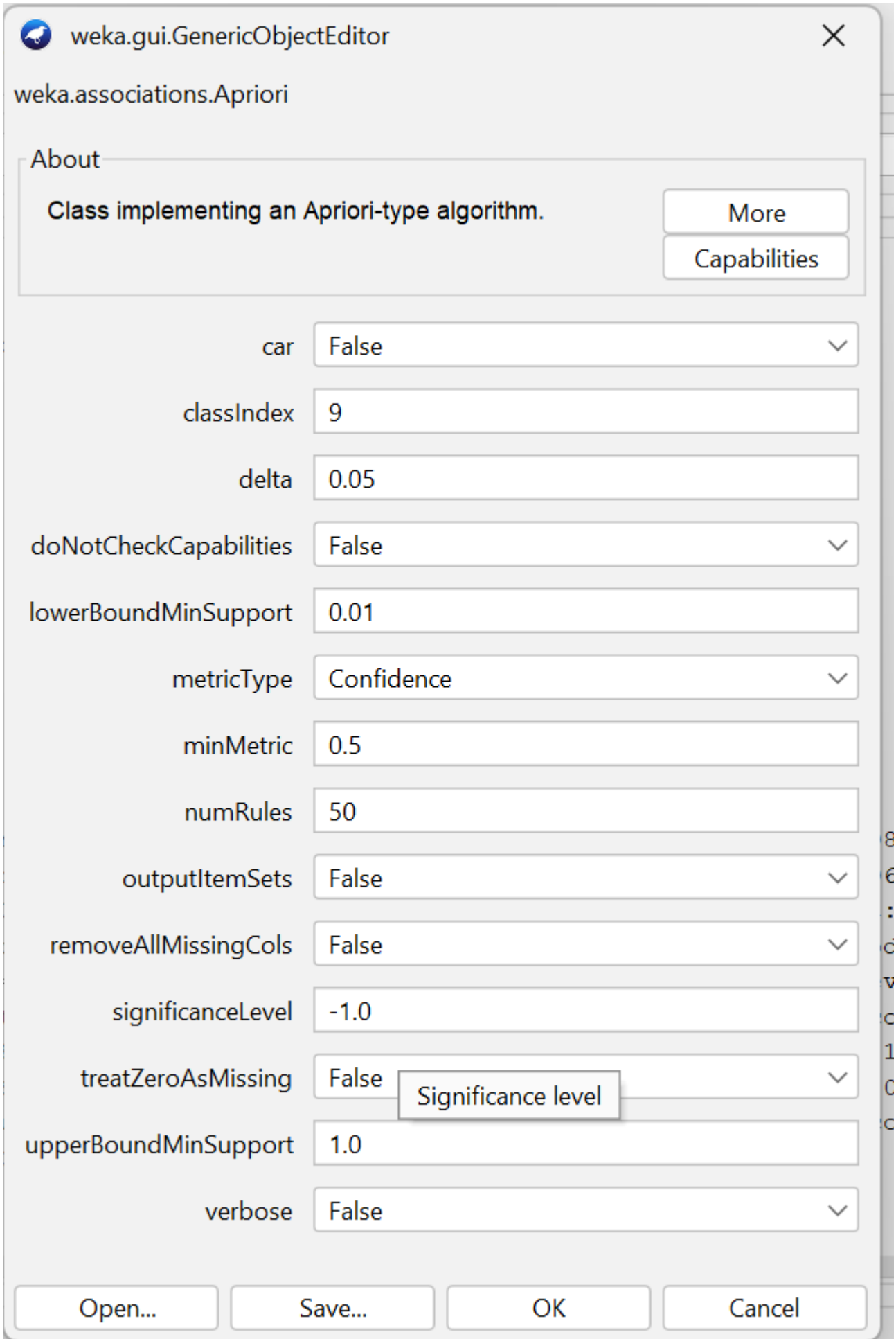
metricType = Confidence

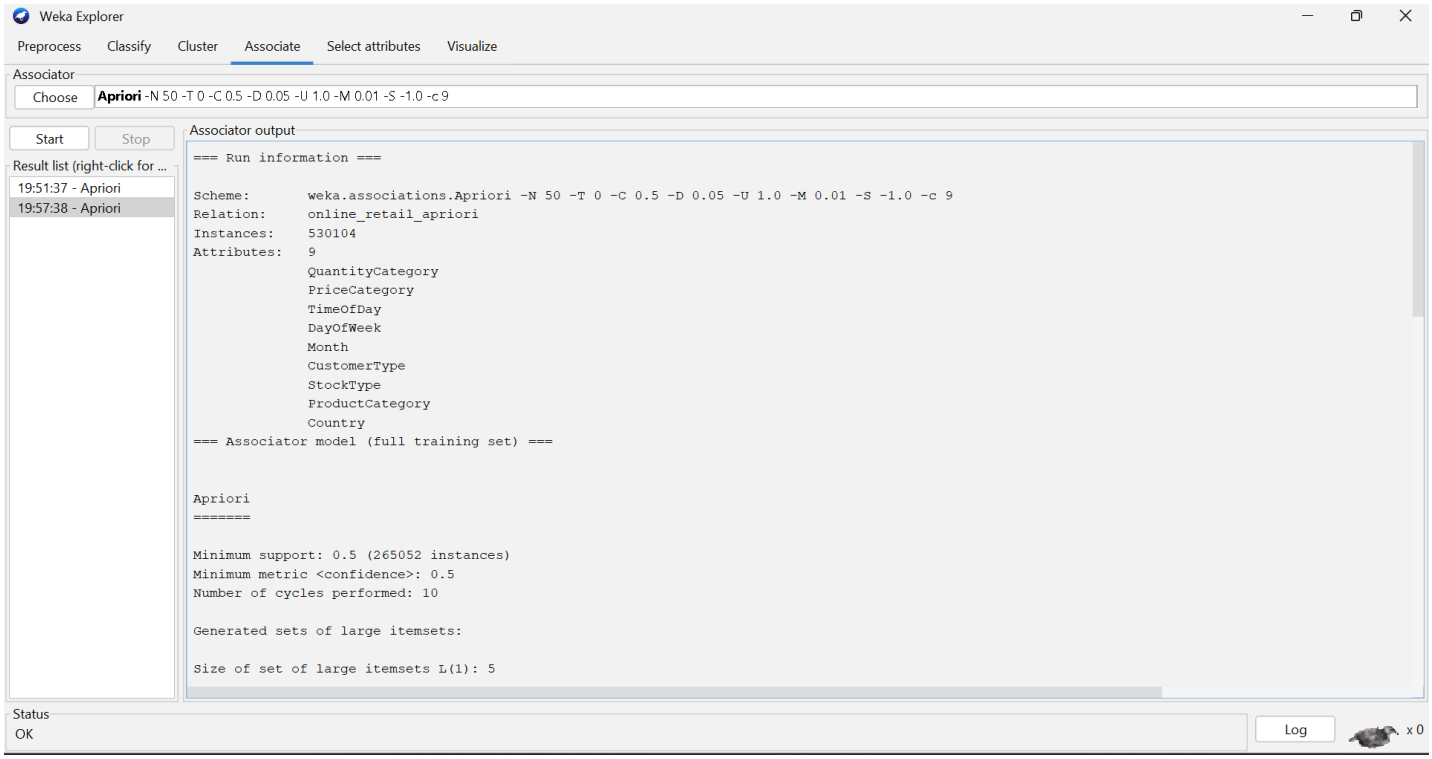
treatZeroAsMissing = false

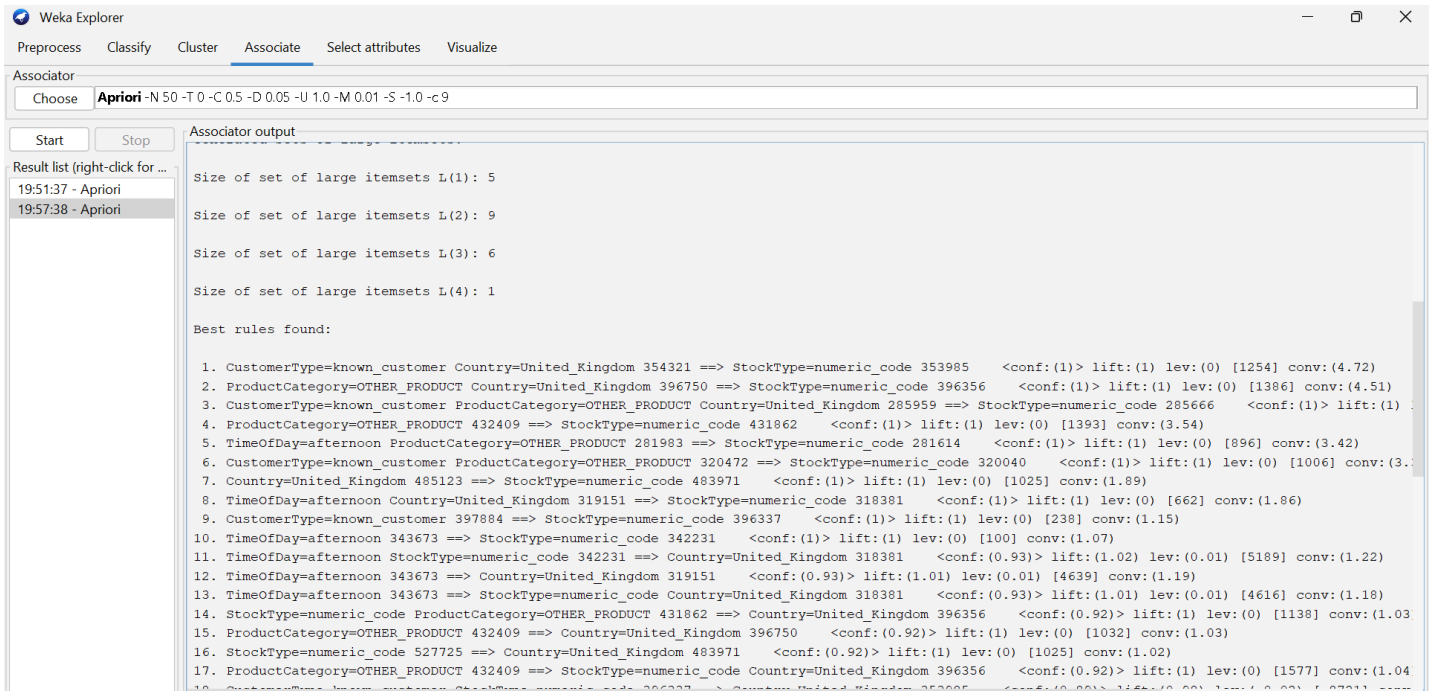
numRules = 50

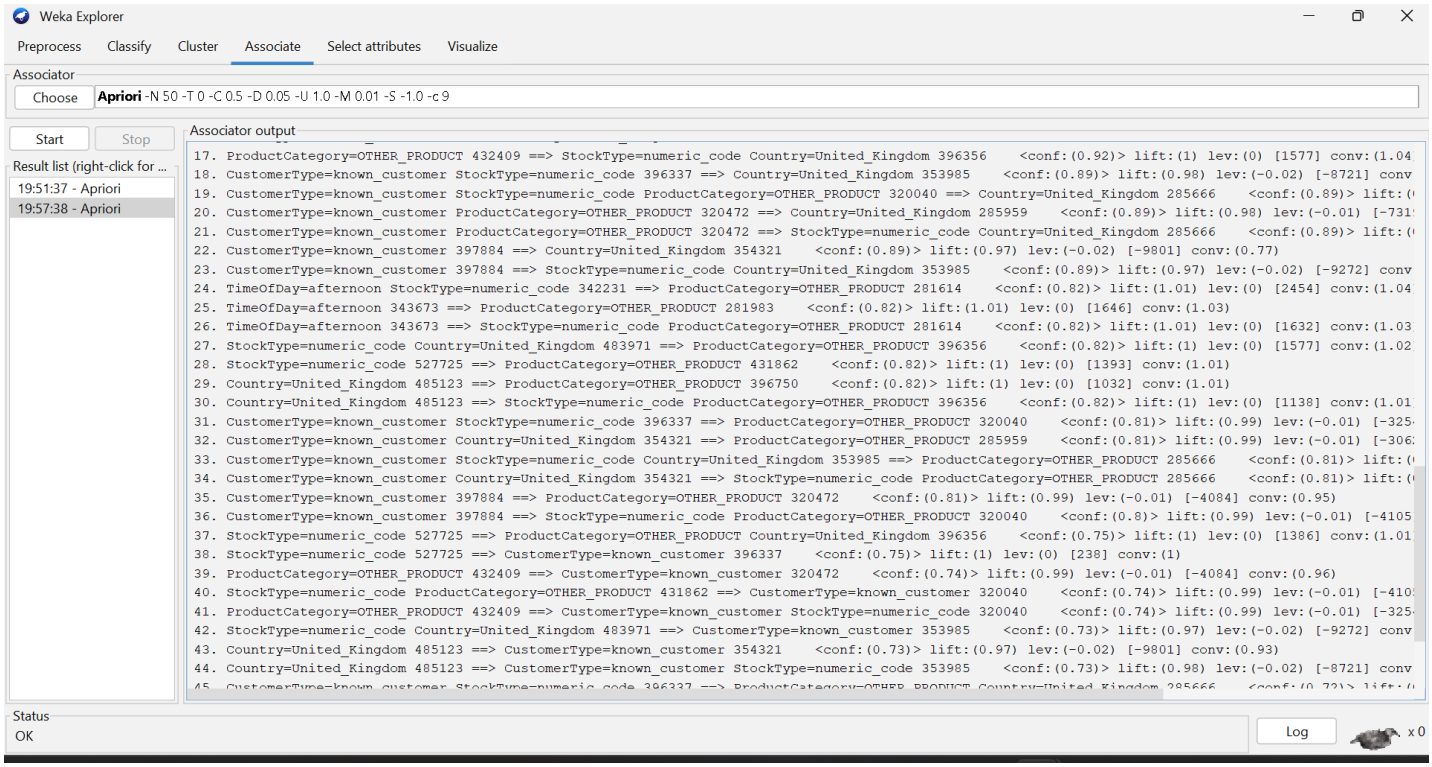
minMetric = 0.5 (50%)

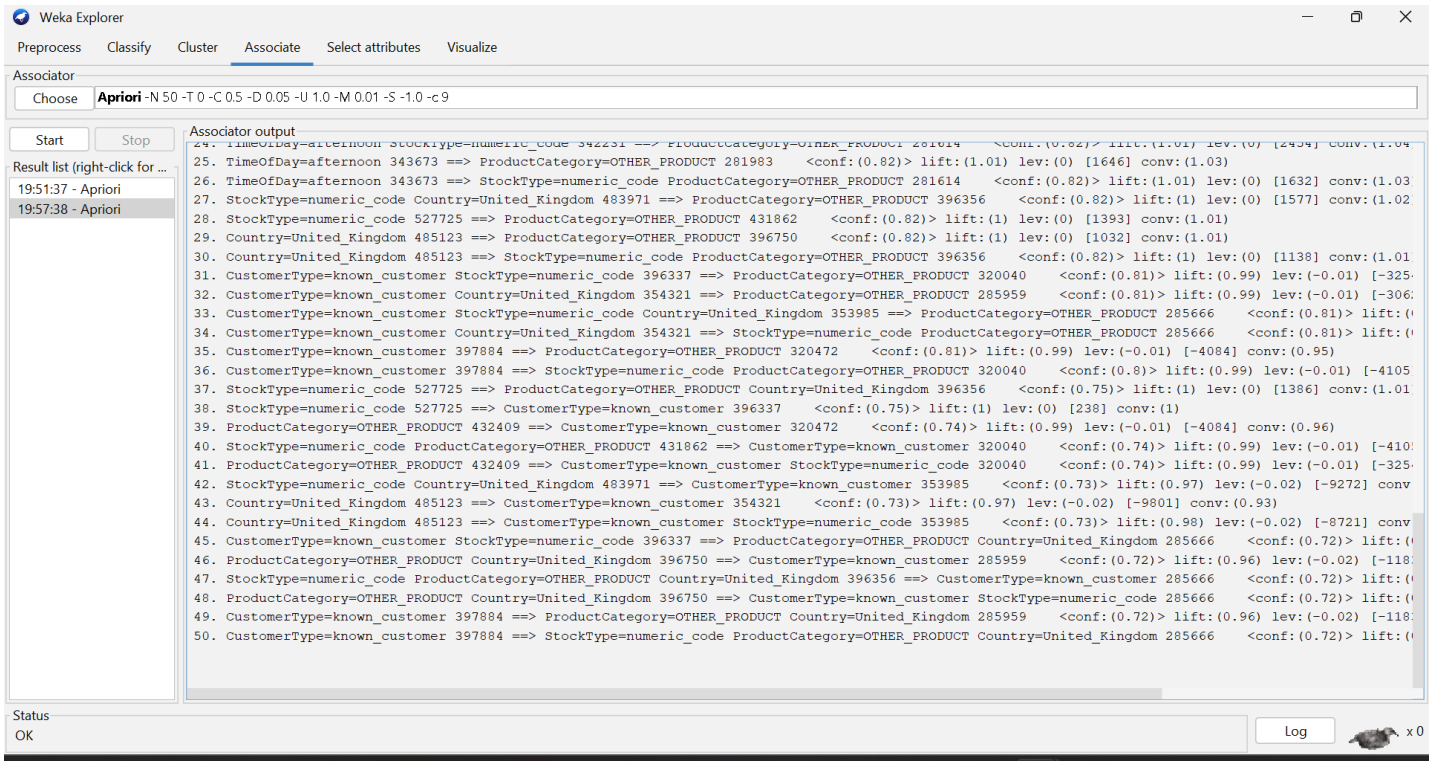
minSupport = 0.01 (1%)


weka.gui.GenericObjectEditor
weka.associations.Apriori
About
Class implementing an Apriori-type algorithm.
More
Capabilities
car
False
classIndex
9
delta
0.05
doNotCheckCapabilities
False
lowerBoundMinSupport
0.01
metricType
Confidence
minMetric
0.5
numRules
50
outputItemSets
False
removeAllMissingCols
False
significanceLevel
-1.0
treatZeroAsMissing
False
Significance level
upperBoundMinSupport
1.0
verbose
False
Open...
Save...
OK
Cancel


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -N 50 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.01 -S -1.0 -c 9
Start
Stop
Result list (right-click for ...
19:51:37 - Apriori
19:57:38 - Apriori
Associator output
=== Run information ===
Scheme: weka.associations.Apriori -N 50 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.01 -S -1.0 -c 9
Relation: online_retail_apriori
Instances: 530104
Attributes: 9
QuantityCategory
PriceCategory
TimeOfDay
DayOfWeek
Month
CustomerType
StockType
ProductCategory
Country
=== Associator model (full training set) ===
Apriori
=======
Minimum support: 0.5 (265052 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 10
Generated sets of large itemsets:
Size of set of large itemsets L(1): 5
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -N 50 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.01 -S -1.0 -c 9
Start
Stop
Result list (right-click for ...
19:51:37 - Apriori
19:57:38 - Apriori
Associator output
Size of set of large itemsets L(1): 5
Size of set of large itemsets L(2): 9
Size of set of large itemsets L(3): 6
Size of set of large itemsets L(4): 1
Best rules found:
1. CustomerType=known_customer Country=United_Kingdom 354321 ==> StockType=numeric_code 353985 <conf:(1)> lift:(1) lev:(0) [1254] conv:(4.72)
2. ProductCategory=OTHER_PRODUCT Country=United_Kingdom 396750 ==> StockType=numeric_code 396356 <conf:(1)> lift:(1) lev:(0) [1386] conv:(4.51)
3. CustomerType=known_customer ProductCategory=OTHER_PRODUCT Country=United_Kingdom 285959 ==> StockType=numeric_code 285666 <conf:(1)> lift:(1)
4. ProductCategory=OTHER_PRODUCT 432409 ==> StockType=numeric_code 431862 <conf:(1)> lift:(1) lev:(0) [1393] conv:(3.54)
5. TimeOfDay=afternoon ProductCategory=OTHER_PRODUCT 281983 ==> StockType=numeric_code 281614 <conf:(1)> lift:(1) lev:(0) [896] conv:(3.42)
6. CustomerType=known_customer ProductCategory=OTHER_PRODUCT 320472 ==> StockType=numeric_code 320040 <conf:(1)> lift:(1) lev:(0) [1006] conv:(3.
7. Country=United_Kingdom 485123 ==> StockType=numeric_code 483971 <conf:(1)> lift:(1) lev:(0) [1025] conv:(1.89)
8. TimeOfDay=afternoon Country=United_Kingdom 319151 ==> StockType=numeric_code 318381 <conf:(1)> lift:(1) lev:(0) [662] conv:(1.86)
9. CustomerType=known_customer 397884 ==> StockType=numeric_code 396337 <conf:(1)> lift:(1) lev:(0) [238] conv:(1.15)
10. TimeOfDay=afternoon 343673 ==> StockType=numeric_code 342231 <conf:(1)> lift:(1) lev:(0) [100] conv:(1.07)
11. TimeOfDay=afternoon StockType=numeric_code 342231 ==> Country=United_Kingdom 318381 <conf:(0.93)> lift:(1.02) lev:(0.01) [5189] conv:(1.22)
12. TimeOfDay=afternoon 343673 ==> Country=United_Kingdom 319151 <conf:(0.93)> lift:(1.01) lev:(0.01) [4639] conv:(1.19)
13. TimeOfDay=afternoon 343673 ==> StockType=numeric_code Country=United_Kingdom 318381 <conf:(0.93)> lift:(1.01) lev:(0.01) [4616] conv:(1.18)
14. StockType=numeric_code ProductCategory=OTHER_PRODUCT 431862 ==> Country=United_Kingdom 396356 <conf:(0.92)> lift:(1) lev:(0) [1138] conv:(1.03
15. ProductCategory=OTHER_PRODUCT 432409 ==> Country=United_Kingdom 396750 <conf:(0.92)> lift:(1) lev:(0) [1032] conv:(1.03)
16. StockType=numeric_code 527725 ==> Country=United_Kingdom 483971 <conf:(0.92)> lift:(1) lev:(0) [1025] conv:(1.02)
17. ProductCategory=OTHER_PRODUCT 432409 ==> StockType=numeric_code Country=United_Kingdom 396356 <conf:(0.92)> lift:(1) lev:(0) [1577] conv:(1.04


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -N 50 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.01 -S -1.0 -c 9
Start
Stop
Result list (right-click for ...
19:51:37 - Apriori
19:57:38 - Apriori
Associator output
17. ProductCategory=OTHER_PRODUCT 432409 ==> StockType=numeric_code Country=United_Kingdom 396356 <conf:(0.92)> lift:(1) lev:(0) [1577] conv:(1.04
18. CustomerType=known_customer StockType=numeric_code 396337 ==> Country=United_Kingdom 353985 <conf:(0.89)> lift:(0.98) lev:(-0.02) [-8721] conv
19. CustomerType=known_customer StockType=numeric_code ProductCategory=OTHER_PRODUCT 320040 ==> Country=United_Kingdom 285666 <conf:(0.89)> lift:(
20. CustomerType=known_customer ProductCategory=OTHER_PRODUCT 320472 ==> Country=United_Kingdom 285959 <conf:(0.89)> lift:(0.98) lev:(-0.01) [-731
21. CustomerType=known_customer ProductCategory=OTHER_PRODUCT 320472 ==> StockType=numeric_code Country=United_Kingdom 285666 <conf:(0.89)> lift:(
22. CustomerType=known_customer 397884 ==> Country=United_Kingdom 354321 <conf:(0.89)> lift:(0.97) lev:(-0.02) [-9801] conv:(0.77)
23. CustomerType=known_customer 397884 ==> StockType=numeric_code Country=United_Kingdom 353985 <conf:(0.89)> lift:(0.97) lev:(-0.02) [-9272] conv
24. TimeOfDay=afternoon StockType=numeric_code 342231 ==> ProductCategory=OTHER_PRODUCT 281614 <conf:(0.82)> lift:(1.01) lev:(0) [2454] conv:(1.04
25. TimeOfDay=afternoon 343673 ==> ProductCategory=OTHER_PRODUCT 281983 <conf:(0.82)> lift:(1.01) lev:(0) [1646] conv:(1.03)
26. TimeOfDay=afternoon 343673 ==> StockType=numeric_code ProductCategory=OTHER_PRODUCT 281614 <conf:(0.82)> lift:(1.01) lev:(0) [1632] conv:(1.03
27. StockType=numeric_code Country=United_Kingdom 483971 ==> ProductCategory=OTHER_PRODUCT 396356 <conf:(0.82)> lift:(1) lev:(0) [1577] conv:(1.02
28. StockType=numeric_code 527725 ==> ProductCategory=OTHER_PRODUCT 431862 <conf:(0.82)> lift:(1) lev:(0) [1393] conv:(1.01)
29. Country=United_Kingdom 485123 ==> ProductCategory=OTHER_PRODUCT 396750 <conf:(0.82)> lift:(1) lev:(0) [1032] conv:(1.01)
30. Country=United_Kingdom 485123 ==> StockType=numeric_code ProductCategory=OTHER_PRODUCT 396356 <conf:(0.82)> lift:(1) lev:(0) [1138] conv:(1.01
31. CustomerType=known_customer StockType=numeric_code 396337 ==> ProductCategory=OTHER_PRODUCT 320040 <conf:(0.81)> lift:(0.99) lev:(-0.01) [-325
32. CustomerType=known_customer Country=United_Kingdom 354321 ==> ProductCategory=OTHER_PRODUCT 285959 <conf:(0.81)> lift:(0.99) lev:(-0.01) [-306
33. CustomerType=known_customer StockType=numeric_code Country=United_Kingdom 353985 ==> ProductCategory=OTHER_PRODUCT 285666 <conf:(0.81)> lift:(
34. CustomerType=known_customer Country=United_Kingdom 354321 ==> StockType=numeric_code ProductCategory=OTHER_PRODUCT 285666 <conf:(0.81)> lift:(
35. CustomerType=known_customer 397884 ==> ProductCategory=OTHER_PRODUCT 320472 <conf:(0.81)> lift:(0.99) lev:(-0.01) [-4084] conv:(0.95)
36. CustomerType=known_customer 397884 ==> StockType=numeric_code ProductCategory=OTHER_PRODUCT 320040 <conf:(0.8)> lift:(0.99) lev:(-0.01) [-4105
37. StockType=numeric_code 527725 ==> ProductCategory=OTHER_PRODUCT Country=United_Kingdom 396356 <conf:(0.75)> lift:(1) lev:(0) [1386] conv:(1.01
38. StockType=numeric_code 527725 ==> CustomerType=known_customer 396337 <conf:(0.75)> lift:(1) lev:(0) [238] conv:(1)
39. ProductCategory=OTHER_PRODUCT 432409 ==> CustomerType=known_customer 320472 <conf:(0.74)> lift:(0.99) lev:(-0.01) [-4084] conv:(0.96)
40. StockType=numeric_code ProductCategory=OTHER_PRODUCT 431862 ==> CustomerType=known_customer 320040 <conf:(0.74)> lift:(0.99) lev:(-0.01) [-410
41. ProductCategory=OTHER_PRODUCT 432409 ==> CustomerType=known_customer StockType=numeric_code 320040 <conf:(0.74)> lift:(0.99) lev:(-0.01) [-325
42. StockType=numeric_code Country=United_Kingdom 483971 ==> CustomerType=known_customer 353985 <conf:(0.73)> lift:(0.97) lev:(-0.02) [-9272] conv
43. Country=United_Kingdom 485123 ==> CustomerType=known_customer 354321 <conf:(0.73)> lift:(0.97) lev:(-0.02) [-9801] conv:(0.93)
44. Country=United_Kingdom 485123 ==> CustomerType=known_customer StockType=numeric_code 353985 <conf:(0.73)> lift:(0.98) lev:(-0.02) [-8721] conv
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -N 50 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.01 -S -1.0 -c 9
Start
Stop
Result list (right-click for ...
19:51:37 - Apriori
19:57:38 - Apriori
Associator output
25. TimeOfDay=afternoon 343673 ==> ProductCategory=OTHER_PRODUCT 281983 <conf:(0.82)> lift:(1.01) lev:(0) [1646] conv:(1.03)
26. TimeOfDay=afternoon 343673 ==> StockType=numeric_code ProductCategory=OTHER_PRODUCT 281614 <conf:(0.82)> lift:(1.01) lev:(0) [1632] conv:(1.03
27. StockType=numeric_code Country=United_Kingdom 483971 ==> ProductCategory=OTHER_PRODUCT 396356 <conf:(0.82)> lift:(1) lev:(0) [1577] conv:(1.02
28. StockType=numeric_code 527725 ==> ProductCategory=OTHER_PRODUCT 431862 <conf:(0.82)> lift:(1) lev:(0) [1393] conv:(1.01)
29. Country=United_Kingdom 485123 ==> ProductCategory=OTHER_PRODUCT 396750 <conf:(0.82)> lift:(1) lev:(0) [1032] conv:(1.01)
30. Country=United_Kingdom 485123 ==> StockType=numeric_code ProductCategory=OTHER_PRODUCT 396356 <conf:(0.82)> lift:(1) lev:(0) [1138] conv:(1.01
31. CustomerType=known_customer StockType=numeric_code 396337 ==> ProductCategory=OTHER_PRODUCT 320040 <conf:(0.81)> lift:(0.99) lev:(-0.01) [-325
32. CustomerType=known_customer Country=United_Kingdom 354321 ==> ProductCategory=OTHER_PRODUCT 285959 <conf:(0.81)> lift:(0.99) lev:(-0.01) [-306
33. CustomerType=known_customer StockType=numeric_code Country=United_Kingdom 353985 ==> ProductCategory=OTHER_PRODUCT 285666 <conf:(0.81)> lift:(
34. CustomerType=known_customer Country=United_Kingdom 354321 ==> StockType=numeric_code ProductCategory=OTHER_PRODUCT 285666 <conf:(0.81)> lift:(
35. CustomerType=known_customer 397884 ==> ProductCategory=OTHER_PRODUCT 320472 <conf:(0.81)> lift:(0.99) lev:(-0.01) [-4084] conv:(0.95)
36. CustomerType=known_customer 397884 ==> StockType=numeric_code ProductCategory=OTHER_PRODUCT 320040 <conf:(0.8)> lift:(0.99) lev:(-0.01) [-4105
37. StockType=numeric_code 527725 ==> ProductCategory=OTHER_PRODUCT Country=United_Kingdom 396356 <conf:(0.75)> lift:(1) lev:(0) [1386] conv:(1.01
38. StockType=numeric_code 527725 ==> CustomerType=known_customer 396337 <conf:(0.75)> lift:(1) lev:(0) [238] conv:(1)
39. ProductCategory=OTHER_PRODUCT 432409 ==> CustomerType=known_customer 320472 <conf:(0.74)> lift:(0.99) lev:(-0.01) [-4084] conv:(0.96)
40. StockType=numeric_code ProductCategory=OTHER_PRODUCT 431862 ==> CustomerType=known_customer 320040 <conf:(0.74)> lift:(0.99) lev:(-0.01) [-410
41. ProductCategory=OTHER_PRODUCT 432409 ==> CustomerType=known_customer StockType=numeric_code 320040 <conf:(0.74)> lift:(0.99) lev:(-0.01) [-325
42. StockType=numeric_code Country=United_Kingdom 483971 ==> CustomerType=known_customer 353985 <conf:(0.73)> lift:(0.97) lev:(-0.02) [-9272] conv
43. Country=United_Kingdom 485123 ==> CustomerType=known_customer 354321 <conf:(0.73)> lift:(0.97) lev:(-0.02) [-9801] conv:(0.93)
44. Country=United_Kingdom 485123 ==> CustomerType=known_customer StockType=numeric_code 353985 <conf:(0.73)> lift:(0.98) lev:(-0.02) [-8721] conv
45. CustomerType=known_customer StockType=numeric_code 396337 ==> ProductCategory=OTHER_PRODUCT Country=United_Kingdom 285666 <conf:(0.72)> lift:(
46. ProductCategory=OTHER_PRODUCT Country=United_Kingdom 396750 ==> CustomerType=known_customer 285959 <conf:(0.72)> lift:(0.96) lev:(-0.02) [-118
47. StockType=numeric_code ProductCategory=OTHER_PRODUCT Country=United_Kingdom 396356 ==> CustomerType=known_customer 285666 <conf:(0.72)> lift:(
48. ProductCategory=OTHER_PRODUCT Country=United_Kingdom 396750 ==> CustomerType=known_customer StockType=numeric_code 285666 <conf:(0.72)> lift:(
49. CustomerType=known_customer 397884 ==> ProductCategory=OTHER_PRODUCT Country=United_Kingdom 285959 <conf:(0.72)> lift:(0.96) lev:(-0.02) [-118
50. CustomerType=known_customer 397884 ==> StockType=numeric_code ProductCategory=OTHER_PRODUCT Country=United_Kingdom 285666 <conf:(0.72)> lift:(
Status
OK
Log
x 0

2.2.4 Result

Apriori Configuration Summary

Two experiments were conducted with different parameter settings. The default run used standard parameters (10% minimum support, 90% minimum confidence, 10 rules) as a baseline. The configured run adjusted the parameters to account for the large dataset size (530,104 instances) and to discover more interesting patterns. Table 4 summarises the key settings and the rationale for each choice in the configured run.

Table: Apriori Parameter Settings and Rationale (Online Retail Dataset)

Parameter	Default Run	Configured Run	Rationale for Configured Run
minSupport	0.1 (10%)	0.01 (1%)	With 530K transactions, 1% = 5,301 instances. Default 10% would require 53,010 instances, which is too strict and would miss many interesting patterns.
minMetric (Confidence)	0.9 (90%)	0.5 (50%)	90% confidence is very strict and may only find trivial rules. 50% provides a lower threshold to discover more varied and potentially interesting patterns.
numRules	10	50	Default 10 rules is too few to explore patterns in a dataset of this size. 50 rules allows broader exploration of the rule space.
metricType	Confidence	Confidence	Confidence is the most interpretable metric for association rules. Lift can be examined later for rules of interest.

treatZeroAsMissing	**false**	**false**	**Our data does not have zeros representing missing values; all attributes are properly labelled nominal values.**
upperBoundMinSupport	**1.0**	**1.0**	**Kept at default.**
delta	**0.05**	**0.05**	**Kept at default.**
outputItemSets	**false**	**false**	**Kept at default (rules only).**
car	**false**	**false**	**Class association rules mode not used; we are exploring general associations, not predicting a specific class.**

Discovered Association Rules – Default Run

The default run (minSupport=0.1, minConfidence=0.9) produced 10 rules, all with confidence ≥ 0.9. However, most rules have lift = 1.0, indicating that the antecedent and consequent are statistically independent – the rule offers no predictive gain over the base probability. The high support threshold (53,010 instances) means only the most common patterns survive. Table 5 lists these rules with their metrics and interpretations.

Table: Association Rules for Online Retail Data (Default Run)

Rule	Antecedent	Consequent	Support (instances)	Support (%)	Confidence	Lift	Interpretation

1	**CustomerType=known_customer AND Country=United_Kingdom**	**StockType=numeric_code**	**354,321**	**66.8%**	**1.0**	**1.0**	**Known UK customers always have a numeric stock code (trivial, as 99.6% of all transactions have numeric codes).**
2	**ProductCategory=OTHER_PRODUCT AND Country=United_Kingdom**	**StockType=numeric_code**	**396,750**	**74.9%**	**1.0**	**1.0**	**Other products bought in the UK always have numeric stock codes.**
3	**ProductCategory=OTHER_PRODUCT**	**StockType=numeric_code**	**432,409**	**81.6%**	**1.0**	**1.0**	**Other products always have numeric stock codes.**
4	**CustomerType=known_customer AND ProductCategory=OTHER_PRODUCT**	**StockType=numeric_code**	**320,472**	**60.5%**	**1.0**	**1.0**	**Known customers buying other products always get numeric stock codes.**
5	**Country=United_Kingdom**	**StockType=numeric_code**	**485,123**	**91.5%**	**1.0**	**1.0**	**Transactions from the UK almost always involve numeric stock codes.**
6	**TimeOfDay=afternoon AND Country=United_Kingdom**	**StockType=numeric_code**	**319,151**	**60.2%**	**1..0**	**1.0**	**Afternoon UK transactions always have numeric stock codes.**
7	**CustomerType=known_customer**	**StockType=numeric_code**	**397,884**	**75.1%**	**1.0**	**1.0**	**Known customers always buy products with numeric stock codes.**

8	**TimeOfDay=afternoon**	**StockType=numeric_code**	**343,673**	**64.8%**	**1.0**	**1.0**	**Afternoon transactions always involve numeric stock codes.**
9	**TimeOfDay=afternoon AND StockType=numeric_code**	**Country=United_Kingdom**	**342,231**	**64.6%**	**0.93**	**1.02**	**Afternoon transactions with numeric stock codes are slightly more likely to be from the UK.**
10	**TimeOfDay=afternoon**	**Country=United_Kingdom**	**343,673**	**64.8%**	**0.93**	**1.01**	**Afternoon transactions are slightly more likely to be from the UK.**

Observations – Default Run:

- **Rules 1-8 all have confidence 1.0 but lift = 1.0, making them trivial – they simply reflect that StockType=numeric_code appears in 99.6% of all transactions.**
- **The only non-trivial rules are 9 and 10, which have lift slightly above 1.0 (1.02 and 1.01). These suggest a very weak tendency for afternoon transactions to be associated with the UK.**
- **The high support threshold (10%) filters out any patterns involving less common attribute values, such as other countries or product categories beyond "OTHER_PRODUCT".**

Discovered Association Rules – Configured Run

The configured run (minSupport=0.01, minConfidence=0.5) produced 50 rules. With support lowered to 1% (5,301 instances), patterns involving less frequent attribute values can now appear. Table 6 presents a selection of the most interesting rules, including those with lift significantly different from 1.0.

Table: Selected Association Rules from Configured Run (Online Retail Data)

Rule	Antecedent	Consequent	Support (instances)	Support (%)	Confidence	Lift	Interpretation
1	CustomerType=known_customer AND Country=United_Kingdom	StockType=numeric_code	354,321	66.8%	1.0	1.0	(Trivial – same as default run)
5	TimeOfDay=afternoon AND ProductCategory=OTHER_PRODUCT	StockType=numeric_code	281,983	53.2%	1.0	1.0	(Trivial)
11	TimeOfDay=afternoon AND StockType=numeric_code	Country=United_Kingdom	342,231	64.6%	0.93	1.02	Weak positive association between afternoon, numeric codes, and UK.

13	**TimeOfDay=afternoon**	**StockType=numeric_code AND Country=United_Kingdom**	**343,673**	**64.8%**	**0.93**	**1.01**	**Afternoon transactions are slightly more likely to involve both numeric codes and UK.**
18	**CustomerType=known_customer AND StockType=numeric_code**	**Country=United_Kingdom**	**396,337**	**74.8%**	**0.89**	**0.98**	**Known customers with numeric codes are slightly *less* likely to be from the UK.**
22	**CustomerType=known_customer**	**Country=United_Kingdom**	**397,884**	**75.1%**	**0.89**	**0.97**	**Known customers are slightly underrepresented in the UK (negative association).**
24	**TimeOfDay=afternoon AND StockType=numeric_code**	**ProductCategory=OTHER_PRODUCT**	**342,231**	**64.6%**	**0.82**	**1.01**	**Weak positive: afternoon numeric-code transactions slightly favour other products.**
29	**Country=United_Kingdom**	**ProductCategory=OTHER_PRODUCT**	**485,123**	**91.5%**	**0.82**	**1.00**	**UK transactions are no more likely to involve other products than average.**

38	StockType=numeric_code	CustomerType=known_customer	527,725	99.6%	0.75	1.00	75% of numeric-code transactions are from known customers – reflects base rate.
41	ProductCategory=OTHER_PRODUCT	CustomerType=known_customer AND StockType=numeric_code	432,409	81.6%	0.74	0.99	Other products are slightly less likely to be bought by known customers with numeric codes.
45	CustomerType=known_customer AND StockType=numeric_code	ProductCategory=OTHER_PRODUCT AND Country=United_Kingdom	396,337	74.8%	0.72	0.96	Known customers with numeric codes are less likely to buy other products in the UK (negative association).
50	CustomerType=known_customer	StockType=numeric_code AND ProductCategory=OTHER_PRODUCT AND Country=United_Kingdom	397,884	75.1%	0.72	0.96	Known customers are less likely to be associated with the combination of numeric code, other product, and UK.

Table: Rules with Highest Lift (Configured Run)

Rule	Antecedent	Consequent	Confidence	Lift
11	TimeOfDay=afternoon AND StockType=numeric_code	Country=United_Kingdom	0.93	1.02

13	TimeOfDay=afternoon	StockType=numeric_code AND Country=United_Kingdom	0.93	1.01
24	TimeOfDay=afternoon AND StockType=numeric_code	ProductCategory=OTHER_PRODUCT	0.82	1.01
9	TimeOfDay=afternoon AND StockType=numeric_code	Country=United_Kingdom (duplicate)	0.93	1.02

Table: Rules with Lowest Lift (Negative Associations)

Rule	Antecedent	Consequent	Confidence	Lift
50	CustomerType=known_customer	StockType=numeric_code AND ProductCategory=OTHER_PRODUCT AND Country=United_Kingdom	0.72	0.96
45	CustomerType=known_customer AND StockType=numeric_code	ProductCategory=OTHER_PRODUCT AND Country=United_Kingdom	0.72	0.96
22	CustomerType=known_customer	Country=United_Kingdom	0.89	0.97

Frequent Itemsets

The frequent itemsets (L1 through L3) reveal the dominant values in the dataset. Table 9 lists itemsets with support ≥ 50% (265,052 instances).

Table: Frequent Itemsets (Support ≥ 50%)

Itemset	Support (instances)	Support (%)

StockType=numeric_code	**527,725**	**99.6%**
Country=United_Kingdom	**485,123**	**91.5%**
ProductCategory=OTHER_PRODUCT	**432,409**	**81.6%**
CustomerType=known_customer	**397,884**	**75.1%**
TimeOfDay=afternoon	**343,673**	**64.8%**
StockType=numeric_code AND Country=United_Kingdom	**483,971**	**91.3%**
ProductCategory=OTHER_PRODUCT AND Country=United_Kingdom	**396,750**	**74.9%**
CustomerType=known_customer AND StockType=numeric_code	**396,337**	**74.8%**
CustomerType=known_customer AND Country=United_Kingdom	**354,321**	**66.8%**
TimeOfDay=afternoon AND StockType=numeric_code	**342,231**	**64.6%**
ProductCategory=OTHER_PRODUCT AND StockType=numeric_code	**431,862**	**81.5%**
TimeOfDay=afternoon AND Country=United_Kingdom	**319,151**	**60.2%**
TimeOfDay=afternoon AND ProductCategory=OTHER_PRODUCT	**281,983**	**53.2%**

Observations

- **Dominance of StockType=numeric_code (99.6%): This attribute value is nearly universal, making any rule with it in the consequent trivial unless lift deviates from 1.0. This explains why the first eight rules in the default run and many in the configured run have lift = 1.0.**
- **UK dominance (91.5%): Similarly, United Kingdom appears in the vast majority of transactions, so rules predicting the UK are also near-trivial unless lift shows a meaningful deviation.**
- **Weak associations only: The highest lift observed in either run is 1.02, indicating extremely weak positive associations. The strongest pattern is a slight tendency for afternoon transactions to be associated with the UK (lift 1.01-1.02).**
- **Negative associations: Several rules involving CustomerType=known_customer have lift slightly below 1.0 (as low as 0.96), suggesting that known customers are marginally less likely to be from the UK or to buy "other products" in the UK. These effects are consistent but very small.**
- **Impact of lowering support: Reducing minSupport from 10% to 1% allowed more itemsets to be considered, but did not reveal any strong novel associations. The additional rules are mostly variations of the same patterns with slightly different antecedents.**
- **Limitations of aggregated data: The absence of strong rules suggests that the discretisation into high-level categories (QuantityCategory, PriceCategory, etc.) may have removed the fine-grained structure needed for meaningful market basket analysis. The original transactional data with individual product codes would likely yield more interesting rules (e.g., "customers who buy product X also buy product Y").**

Comparison with Nursery Dataset

Aspect	Nursery Dataset	Online Retail Dataset
Strong associations found?	Yes – multiple rules with lift = 3.0	No – maximum lift = 1.02
Trivial rules	Few – most rules provided real insight	Many – due to dominant attribute values
Effect of parameter tuning	Revealed class-specific patterns	Increased rule count but not rule quality
Data granularity	Original categorical values	Highly aggregated categories

References

University of California, Irvine. (n.d.). UCI Machine Learning Repository. http://archive.ics.uci.edu/ml

Chen, D. (2015). Online Retail [Data set]. UCI Machine Learning Repository. https://archive.ics.uci.edu/dataset/352/online+retail

Bohanec, M., & Rajkovic, V. (1997). Nursery [Data set]. UCI Machine Learning Repository. https://archive.ics.uci.edu/dataset/76/nursery