

---

# **Data Mining**

## **SPSS Clementine 12.0**

### **5. Apriori Algorithm**

**Spring 2010**  
Instructor: Dr. Masoud Yaghini

# Outline

---

- **Overview**
- **Apriori Node**
- **Example of Apriori Node: Market Basket Analysis**
- **References**

---

# Overview

# Overview

---

- **Association rules** associate a particular conclusion (the purchase of a particular product) with a set of conditions (the purchase of several other products).
- For example, the rule  
**beer <= cannedveg & frozenmeal (173, 17.0%, 0.84)**
  - states that *beer* often occurs when *cannedveg* and *frozenmeal* occur together.
  - The rule is 84% reliable and applies to 17% of the data, or 173 records.

# Overview

---

- Association rule algorithms are supported:
- Generalized Rule Induction (GRI)
  - discovers association rules in the data.
  - For example, customers who purchase razors and aftershave lotion are also likely to purchase shaving cream.
  - GRI extracts rules with the highest information content based on an index that takes both the generality (support) and accuracy (confidence) of rules into account.
  - GRI can handle numeric and categorical inputs, but the target must be categorical.
- Apriori model
  - extracts a set of rules from the data, pulling out the rules with the highest information content.
  - Apriori requires that input and output fields all be categorical.

# Overview

---

- **CARMA model**

- The CARMA model offers build settings for rule support (support for both antecedent and consequent) rather than just antecedent support.

- **Sequence model**

- discovers association rules in sequential or time-oriented data.
- A sequence is a list of item sets that tends to occur in a predictable order. For example, a customer who purchases a razor and aftershave lotion may purchase shaving cream the next time he shops.

# Overview

---

- Data used by association rule models may be in
  - Transactional format
  - Tabular format

# Overview

## ● Transactional Format

- Transactional data have a separate record for each transaction or item.
- If a customer makes multiple purchases, for example, each would be a separate record, with associated items linked by a customer ID.
- This is also sometimes known as **till-roll format**.

Customer	Purchase
1	jam
2	milk
3	jam
3	bread
4	jam
4	bread
4	milk

- The Apriori, CARMA, and Sequence nodes can all use transactional data.



# Overview

- **Tabular Data**

- Tabular data (also known as **basket** or **truth-table** data) have items represented by separate flags, where each flag field represents the presence or absence of a specific item.
- Each record represents a complete set of associated items.
- Flag fields can be **categorical** or **numeric**, although certain models may have more specific requirements.

Customer	Jam	Bread	Milk
1	T	F	F
2	F	F	T
3	T	T	F
4	T	T	T

- The GRI, Apriori, CARMA,, and Sequence nodes can all use tabular data.

---

# Apriori Node

# Apriori Node

---

- The **Apriori** node is available with the **Association** module.
- This node discovers association rules in the data.
- **Apriori** offers five different methods of selecting rules and uses a sophisticated indexing scheme to efficiently process large datasets.

# Apriori Node

---

- Requirements

- To create an Apriori ruleset, you need one or more *In* fields and one or more *Out* fields.
- Input and output fields (those with direction *In*, *Out*, or *Both*) must be **symbolic**.
- Fields with direction *None* are ignored.
- Fields types must be fully instantiated before executing the node.
- Data can be in **tabular** or **transactional** format.

---

# **Example of Apriori Node: Market Basket Analysis**

# Market Basket Analysis

---

- Using a **Variable File** node, connect to the dataset **BASKETS1n**, selecting to read field names from the file.
- Connect a **Type** node to the data source, and then connect the node to a **Table** node.
- Set the type of the field *cardid* to *Typeless* (because each loyalty card ID occurs only once in the dataset and can therefore be of no use in modeling).
- Select **Set** as the type for the field *sex*
  - this is to ensure that the **Apriori** algorithm will not treat *sex* as a flag

# Market Basket Analysis

---

- Now execute the stream to instantiate the **Type** node and display the table.
- The dataset contains 18 fields, with each record representing a basket.
- **The 18 fields are presented in the following headings:**
  - **Basket summary:**
    - ◆ *cardid*. Loyalty card identifier for customer purchasing this basket.
    - ◆ *value*. Total purchase price of basket.
    - ◆ *pmethod*. Method of payment for basket.

# Market Basket Analysis

---

- The 18 fields are presented in the following headings (cont.):
  - Personal details of cardholder:
    - ◆ *sex*
    - ◆ *homeown*. Whether or not cardholder is a homeowner.
    - ◆ *Income*
    - ◆ *age*
  - Basket contents—flags for presence of product categories:
    - ◆ *fruitveg* , *freshmeat*, *dairy*, *cannedveg*, *cannedmeat*, *frozenmeal*, *beer*, *wine*, *softdrink*, *fish*, *confectionery*



# Market Basket Analysis

---

- **Discovering associations in Basket Contents:**
- Select the fields to be used in this modeling process by editing the **Type** node and setting the directions of all of the product categories to **Both** and all other directions to **None**.
  - You can set options for multiple fields using Shift-click to select the fields before specifying an option from the columns.

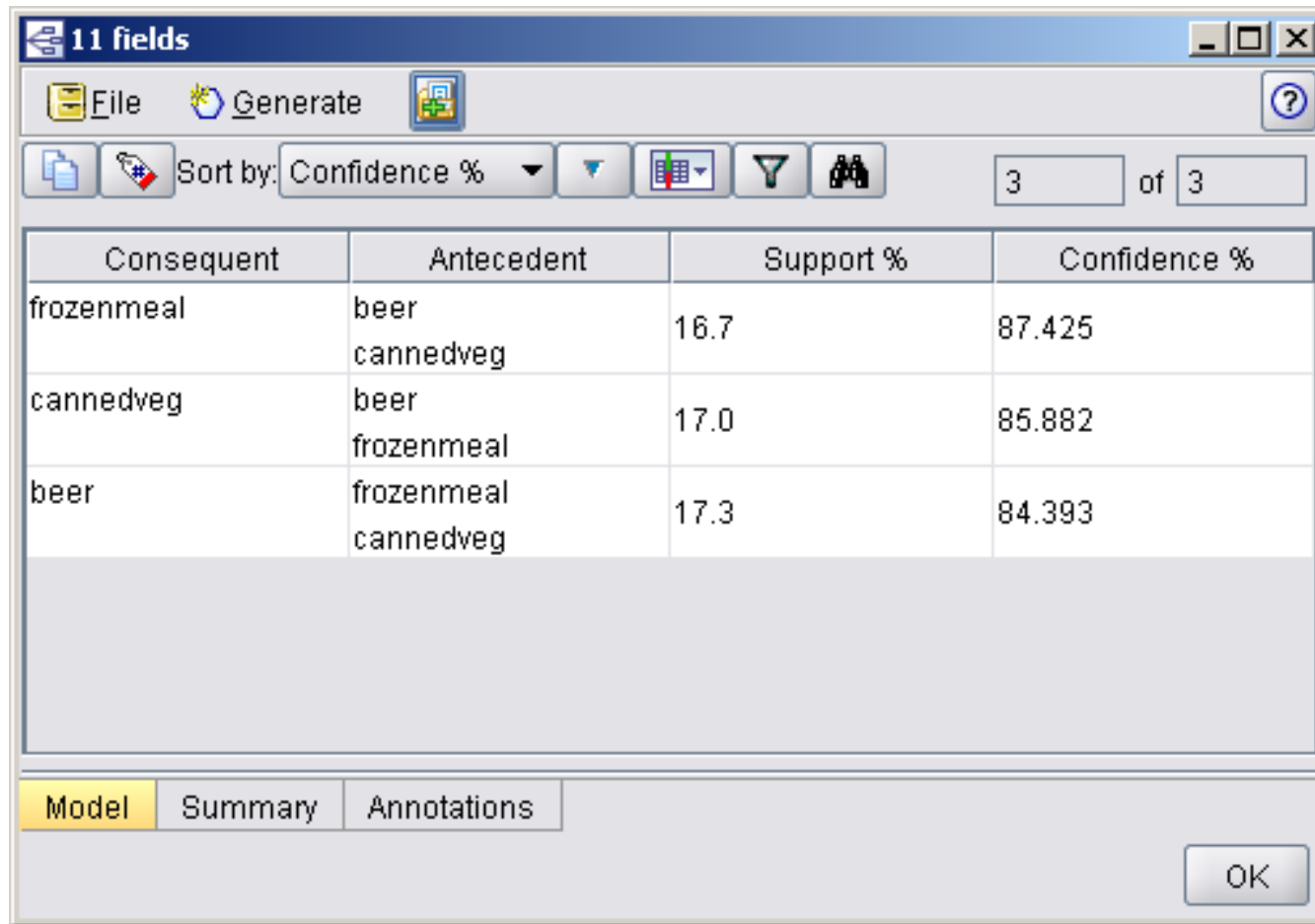
# Market Basket Analysis

---

- Once you have specified fields for modeling, attach a **Apriori** node to the **Type** node
- Select the options, and execute the **Apriori** node.
- The result, an unrefined model on the **Models** tab at the upper right of the managers window, contains association rules that you can view by using the context menu and selecting **Browse**.

# Market Basket Analysis

- These rules show a variety of associations between products



Consequent	Antecedent	Support %	Confidence %
frozenmeal	beer cannedveg	16.7	87.425
cannedveg	beer frozenmeal	17.0	85.882
beer	frozenmeal cannedveg	17.3	84.393

---

# References

# References

---

- Integral Solutions Limited., **Clementine® 12.0 Source, Process, and Output Nodes**, 2007. (Chapter 2)



The end