



CST8390 - Business Intelligence and Data Analytics

Lab 9 - Association Rule

Name: - Id:

Due Date: Week 12 in corresponding lab sessions.

Introduction

The goal of this lab is to perform **Association Rule Mining** on **Super Market** dataset.

Steps

1. Open **Weka** and load the file **supermarket.arff** from “data” directory of **Weka**.
2. From the **preprocess** tab, click on the **Edit** button to view the instances. The “t” letters show which items were purchased.
3. Close Edit window and look at the **attributes**.
 - a. Number of attributes: .
 - b. Number of instances: .
4. Find **tea**, **coffee**, **medicines** and **flowers** and see **how many times** each of the item was purchased?
Tea:
Coffee:
Medicines:
Flowers:
5. Click on the “**Associate**” tab. The **Apriori algorithm** should already be selected but click on the **text field** to **edit** the parameters. Find the **lowerBoundMinSupport**. This is the minimum support percentage that is required to create the rule sets. Set it to **0.25** (i.e., 25%). Set the “**numRules**” to **15**, to print out the **top 15 rules** that are found. Click “**Ok**” to close the window and then click “**Start**” to run the algorithm.

6. The algorithm should run for a number of seconds and then return with **no rules**. That means that no rules were found that have a minimum support of **20%**. Lower the support to **15%** and run it again. Set **numRules** to **50**. How many **rules** were generated this time? .

7. The rules are **sorted** from highest lift to lowest. The lift tells you **how often** the rules are related, or the strength of the rule. Which rules have the **highest lift**?

8. Lower the support now to **10%** and re-run the algorithm. Since more rules are included in the search, this time it should take a long time to run. What is the **highest lift** now that was found and what are the rules?

Example:

If you get:

*frozen foods=t fruit=t total=high 969 ==> bread and cake=t 877 <conf:(0.91)>
lift:(1.26) lev:(0.04) [179] conv:(2.92),*

you need to write the rule as:

frozen foods, fruit ==> bread and cake (conf: 0.91, lift: 1.26).

REMEMBER:

In order to get grades, you need to upload filled-in answer document and screenshots from steps 6-8.

FOR YOUR ANALYSIS:

* **Option 1:** Use your own words to explain **Association Rule Mining** and situations where apply it.

* **Option 2:** What is the strategy to identify association rules in a specific scenario?