CS 634 DATA MINING MIDTERM PROJECT

SUNDAY DAVID

03/19/2021

# PY\_APRIORI MODULE

This is a basic implementation of the Apriori Algorithm in Python

Github link: <https://github.com/sdavid501/Apriori-Algorithm>

## Requirements

* macOS/Windows
* Python 3.6

## Setup

Download and extract the zip file (DAVID\_SUNDAY\_MIDTERMPROJ.zip) into a directory that you’ll use. For illustration purpose, my extracted directory is: /Users/sdavid/Downloads/project

Change your directory to DAVID\_SUNDAY\_MIDTERMPROJ

Text

Description automatically generated

**Setup Python environment**

* You can skip this section if python is already installed in your machine
* Go to <https://www.python.org/downloads/>
* Follow the instruction for installing python 3.6

**Support & Confidence Algorithm**

Someone buys X, also buys Z, X -> Z

Support:

* The number of transactions that include both items (X & Z) divided by the total number of transactions

Confidence:

* The Support of (X & Z) divided by the Support of X

**Input Structure**

The algorithm expects two types of csv files, the first is a csv file of the set of all transaction items, and the second is a csv file of all transactions. The set of items and the transactions are expected as the last column of the csv files.

**Running**

To run the algorithm, run python DAVID\_SUNDAY\_MIDTERMPROJ.py. There are 7 databases in the model, and you’ll be prompted to select one. You’ll also be prompted to enter minimum support (in %) then confirm the value by entering ‘Y’ or ‘y’. In addition, you’ll enter the confidence (in %) then confirm the value by entering ‘Y’ or ‘y’.

Text

Description automatically generated

**Output**

The output of the algorithm is given just after entering the required minimum support and confidence. For illustration, the result below shows the Frequent itemsets and all association rules of Amazon’s database using minimum support of 40% and confidence of 60%.

**A picture containing text, green

Description automatically generated**