# Apriori (Market Basket Analysis)

One of the most important strategies for identifying associations between items is market basket analysis. It operates by searching the data for groups of entries that commonly appear together.

In order to uncover strong rules found in the data using measures of interestingness, association rules are frequently employed to analyse retail basket or transaction data.

***For example, let us assume there are 100 customers.***

10 of them bought milk, 8 bought bread and 6 bought both.

bought milk => bought bread

***support*** = P (Milk & Bread) = 6/100 = 0.06

***confidence*** = support/P(Bread) = 0.06/0.08 = 0.75

***lift*** = confidence/P(Milk) = 0.75/0.10 = 7.5

### *Implementation of Apriori algorithm in health care domain*

Circ Res. 2017 Oct 13;121(9):1092-1101. doi: 10.1161/CIRCRESAHA.117.311312.

J Healthc Eng. 2021 Nov 1;2021:9018408. doi: 10.1155/2021/9018408.

Front Public Health. 2021 Aug 30;9:707062. doi: 10.3389/fpubh.2021.707062.

Risk Manag Healthc Policy. 2021 Jul 12;14:2953-2963. doi: 10.2147/RMHP.S316361.

# Sequential pattern mining (SPM)- SPADE

Market Basket Analysis (MBA) answers products/events that often purchased/occur together (e.g., milk and bread).

Assume that, instead, we want to analyse how a customer's services have changed over **time**, such as figuring out whether purchasing milk in the past predicts a higher possibility of purchasing bread in the future.

To do this, we employ a sequential variation of MBA, also known as "sequential itemset mining" or "sequential pattern mining," to add a time component to the analysis.

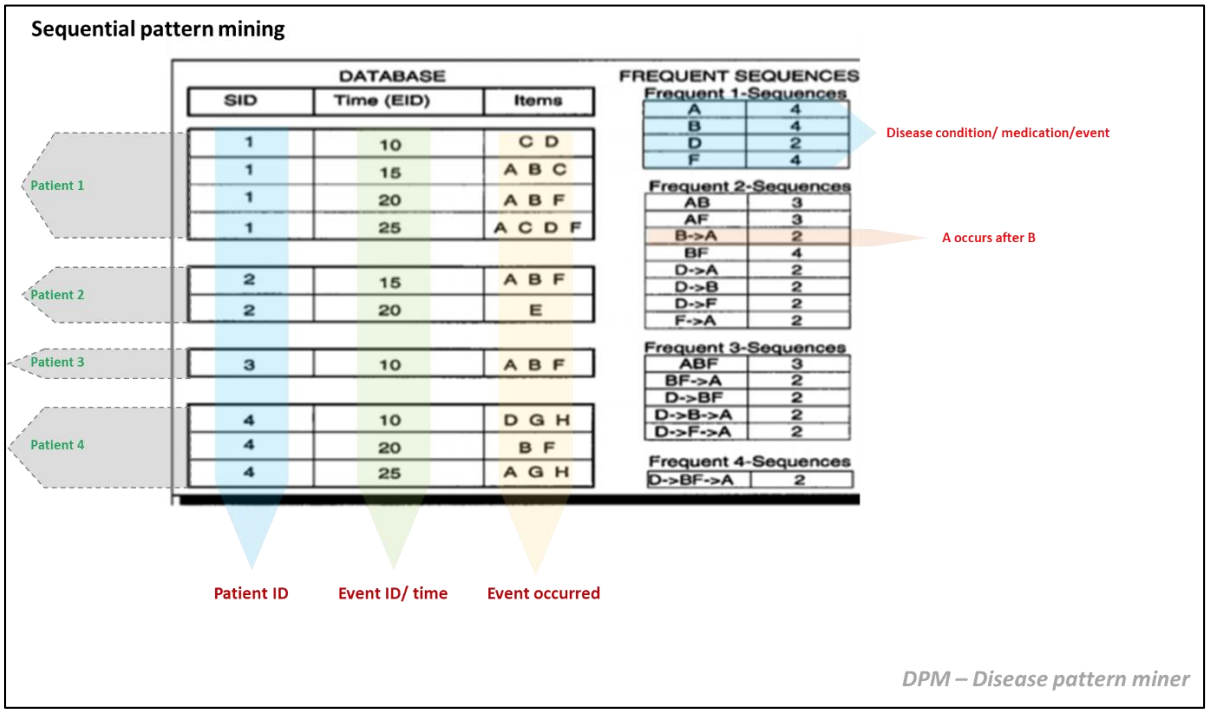


*Image: J. Zaki. (2001). SPADE: An Efficient Algorithm for Mining Frequent Sequences. Machine Learning Journal, 42, 31–60.*

## *Implementation of cSpade algorithm in health care domain*

PLoS One. 2018 Apr 26;13(4):e0194407. doi: 10.1371/journal.pone.0194407.

J Biomed Inform. 2008 Feb;41(1):165-79. doi: 10.1016/j.jbi.2007.05.004.

J Biomed Inform. 2015 Feb;53:73-80. doi: 10.1016/j.jbi.2014.09.003.

Int J Health Plann Manage. 2022 Sep;37(5):2569-2584. doi: 10.1002/hpm.3483.

Conf Proc IEEE Eng Med Biol Soc. 2006;2006:5814-7. doi: 10.1109/IEMBS.2006.

## SERENDIP

Sequential rule mining is a well-established data mining technique for binary valued data. Many variations have been proposed, most approaches use the support-confidence-lift framework. Existing approaches make assumptions concerning the definition of what a sequence is. However, this definition is application dependent.

Sequential rule mining with respect to long-term disease prediction, which entails a rethink of the definition of what a sequence is, and a consequent rethink of the operation of the support-confidence-lift framework. A novel sequential rule mining algorithm is proposed designed to address the challenge of long-term disease prediction. The SEquential RElational N-DIsease Pattern (SERENDIP) algorithm.

Traditional sequential pattern mining approaches assume a sequence is any set of items that occur in order, which may be preceded and/or proceeded with other items or be interrupted by other items. SERENDIP accounts for order of events.

| Example Sequential Patterns | Retail Exemplar Scenario | Multi-morbidity Disease Prediction Scenario |
|---|---|---|
|  | ✓ | ✓ |
|  | ✓ | X |
|  | ✓ | X |
|  | ✓ | X |