

Khai thác Tập luật phổ biến

Giảng viên: Tiến sĩ Bùi Thanh Hùng
Email: tg_buithanhhung@tdtu.edu.vn
Website: <https://sites.google.com/site/hungthanhbui1980>

Online Retail Data Set

Abstract: This is a transnational data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail.

Data Set Characteristics:	Multivariate, Sequential, Time-Series	Number of Instances:	541909	Area:	Business
Attribute Characteristics:	Integer, Real	Number of Attributes:	8	Date Donated	2015-11-06
Associated Tasks:	Classification, Clustering	Missing Values?	N/A	Number of Web Hits:	650424

Attribute Information:

- ✓ InvoiceNo: Invoice number. Nominal, a 6-digit integral number uniquely assigned to each transaction. If this code starts with letter 'c', it indicates a cancellation.
- ✓ StockCode: Product (item) code. Nominal, a 5-digit integral number uniquely assigned to each distinct product.
- ✓ Description: Product (item) name. Nominal.
- ✓ Quantity: The quantities of each product (item) per transaction. Numeric.
- ✓ InvoiceDate: Invoice Date and time. Numeric, the day and time when each transaction was generated.
- ✓ UnitPrice: Unit price. Numeric, Product price per unit in sterling.
- CustomerID: Customer number. Nominal, a 5-digit integral number uniquely assigned to each customer.
- ✓ Country: Country name. Nominal, the name of the country where each customer resides.

Download: <http://archive.ics.uci.edu/ml/datasets/Online+Retail#>

Hãy tìm tập luật phổ biến với các giao dịch ở France, United Kingdom, Portugal, Sweden với các yêu cầu sau:

- ✓ Áp dụng giải thuật Apriori, FPGrowth cho mỗi nước với tham số $\text{min_support} = 0.05$
- ✓ Áp dụng `association_rules` với tham số `metric = "lift"` và `min_threshold = 1`
- ✓ Sắp xếp các luật theo 'confidence' và 'lift' và in ra kết quả
- ✓ Nhận xét kết quả đạt được ở mỗi nước