Data Mining for Business

Association Mining Techniques

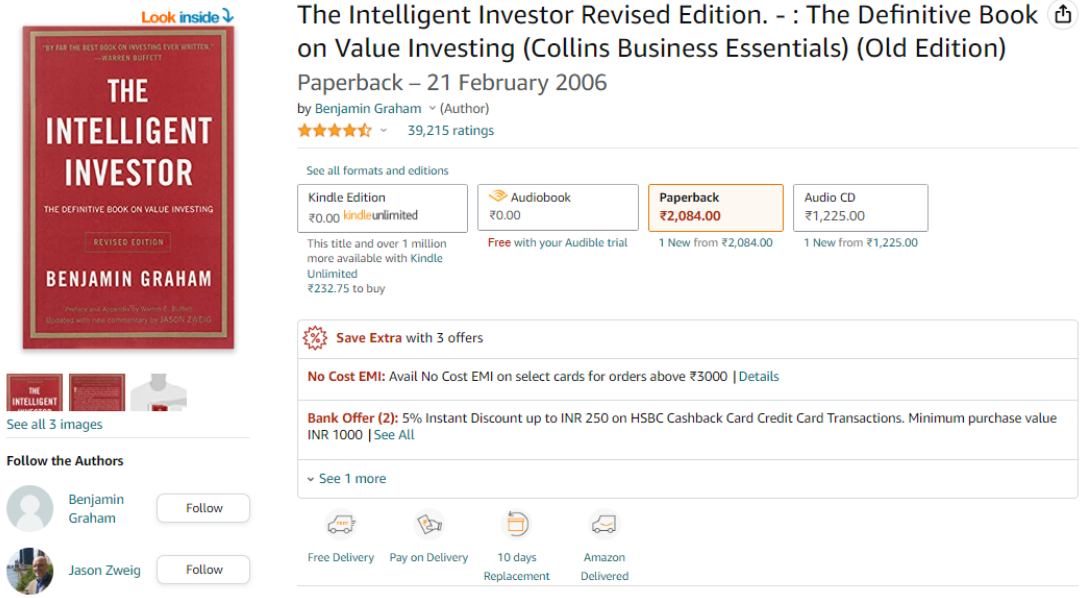
Dr. Shipra Maurya
Department of Management Studies
IIT (ISM) Dhanbad
Email: shipra@iitism.ac.in

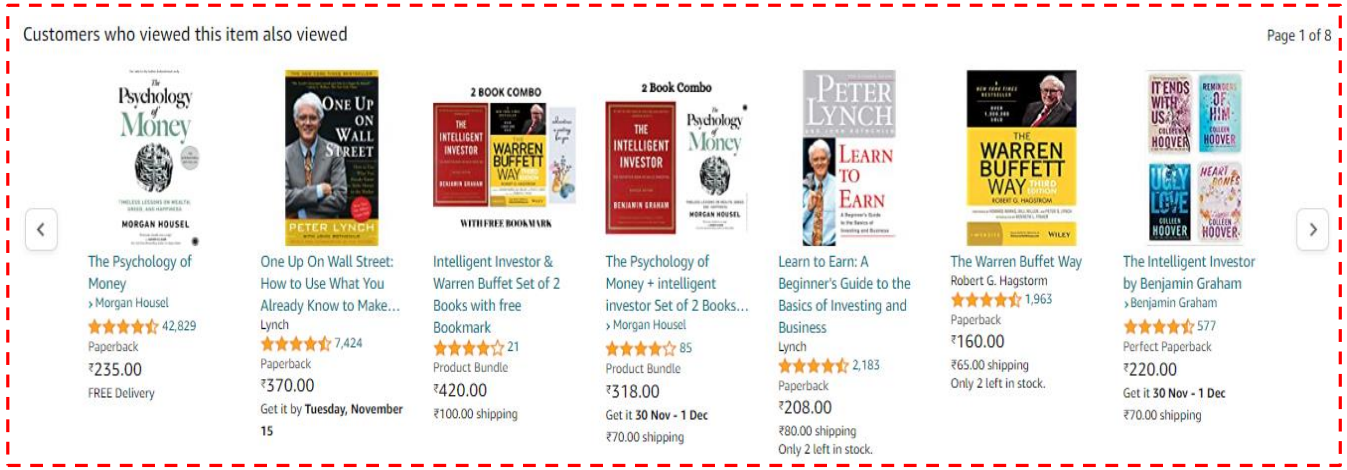
Association Mining

- Association Mining/ Association Rules/ Affinity analysis/ Market Basket analysis
- Aim is to identify groups of products tend to be purchased together. It is an unsupervised learning method
- study of "what goes with what"
- These items can then be displayed together, offered in post transaction coupons, or recommended in online shopping

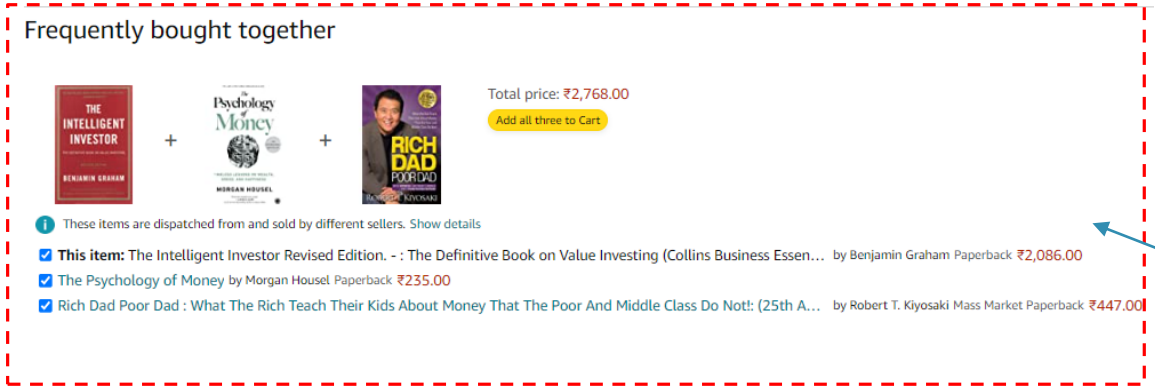
Association Mining

- Association rules are commonly encountered in online recommendation systems (or recommender systems)





Based on Collaborative filtering



Based on Association Rules

Apriori Algorithm

- The key idea of the algorithm is to begin by generating frequent itemsets with just one item (one-itemsets) and to recursively generate frequent itemsets with two items, then with three items, and so on, until we have generated frequent itemsets of all sizes

$$\text{Confidence} = \frac{\text{no. of transactions with both antecedent and consequent itemsets}}{\text{no. of transactions with antecedent itemset}}.$$

$$\text{Support} = \hat{P}(\text{antecedent AND consequent})$$ OR Number of transactions with both antecedent and consequent / number of transaction in database

$$\text{lift ratio} = \frac{\text{confidence}}{\text{benchmark confidence}}.$$

$$\text{Benchmark confidence} = \frac{\text{no. of transactions with consequent itemset}}{\text{no. of transactions in database}}.$$

Apriori Algorithm

- Support for the rule indicates its impact in terms of overall size: How many transactions are affected?
- Lift ratio indicates how efficient the rule is in finding consequents, compared to random selection. A lift ratio >1 suggest the rule is useful
- Confidence indicates at what rate consequents will be found and is useful in identifying the business usefulness of a rule.

Apriori Algorithm Steps

- **Step 1** – Identify all frequent itemsets i.e. remove item combinations that are rare in dataset
- **Step 2** – From the frequent itemsets, generate association rules that meet a confidence requirement i.e. filter the remaining rules and select only those with high confidence

Output – The output of the association analysis will be identified association rules that will be further used by the organization in making business decisions such as store layout, item placement, item to be recommended in online shopping etc.

Association Mining Example

Transaction	Faceplate colors purchased			
1	red	white	green	
2	white	orange		
3	white	blue		
4	red	white	orange	
5	red	blue		
6	white	blue		
7	red	blue		
8	red	white	blue	green
9	red	white	blue	
10	yellow			

Transaction	Red	White	Blue	Orange	Green	Yellow
1	1	1	0	0	1	0
2	0	1	0	1	0	0
3	0	1	1	0	0	0
4	1	1	0	1	0	0
5	1	0	1	0	0	0
6	0	1	1	0	0	0
7	1	0	1	0	0	0
8	1	1	1	0	1	0
9	1	1	1	0	0	0
10	0	0	0	0	0	1

Association Mining Example

Transaction	Red	White	Blue	Orange	Green	Yellow
1	1	1	0	0	1	0
2	0	1	0	1	0	0
3	0	1	1	0	0	0
4	1	1	0	1	0	0
5	1	0	1	0	0	0
6	0	1	1	0	0	0
7	1	0	1	0	0	0
8	1	1	1	0	1	0
9	1	1	1	0	0	0
10	0	0	0	0	0	1

Itemset	Support (count)
{red}	6
{white}	7
{blue}	6
{orange}	2
{green}	2
{red, white}	4
{red, blue}	4
{red, green}	2
{white, blue}	4
{white, orange}	2
{white, green}	2
{red, white, blue}	2
{red, white, green}	2

Rule	Confidence	Lift
$\{red, white\} \Rightarrow \{green\}$	$\frac{\text{support of } \{red, white, green\}}{\text{support of } \{red, white\}} = \frac{2}{4} = 50\%$	$\frac{\text{confidence of rule}}{\text{benchmark confidence}} = \frac{50\%}{20\%} = 2.5$
$\{green\} \Rightarrow \{red\}$	$\frac{\text{support of } \{green, red\}}{\text{support of } \{green\}} = \frac{2}{2} = 100\%$	$\frac{\text{confidence of rule}}{\text{benchmark confidence}} = \frac{100\%}{60\%} = 1.67$
$\{white, green\} \Rightarrow \{red\}$	$\frac{\text{support of } \{white, green, red\}}{\text{support of } \{white, green\}} = \frac{2}{2} = 100\%$	$\frac{\text{confidence of rule}}{\text{benchmark confidence}} = \frac{100\%}{60\%} = 1.67$

Interpretation – If white and green are purchased then with confidence 100%, red will also be purchased. This rule has lift ratio of 1.67.

Thank you!

You can reach me on :

Email : shipra@iitism.ac.in

LinkedIn : https://www.linkedin.com/in/shipra-maurya1205/?originalSubdomain=in