# DMBI Experiment 8 – Association Rule Mining using Rapid Miner Tool

Name: Atharva Lotankar
Class: D15C;   **Rnum:** 27
Batch B

Aim: To implement Association Rule Mining (Apriori Algorithm) using RapidMiner tool.

Theory:

Association rule mining is a fundamental technique in data mining used to uncover hidden relationships, patterns, and correlations within large datasets. It helps organizations identify frequent itemsets and derive meaningful rules in the form of "if-then" statements, such as products often purchased together. This method is widely applied in market basket analysis, recommendation systems, fraud detection, and cross-selling strategies, making it valuable for decision-making and business intelligence.
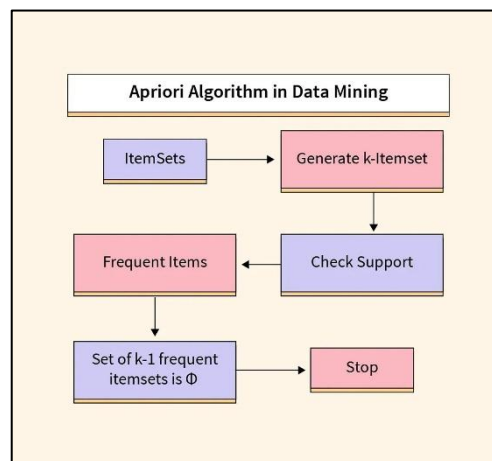
The Apriori algorithm is one of the most popular approaches for association rule mining. It works by identifying frequent itemsets based on minimum support thresholds and then generating strong association rules using confidence and lift measures. Its iterative, bottom-up nature ensures efficiency by pruning unpromising itemsets early, saving both time and computational resources. Despite its simplicity, Apriori remains a strong choice for deriving reliable insights from transactional and categorical data.

RapidMiner provides a powerful platform to implement association rule mining and Apriori efficiently. With its visual workflow design, users can load datasets, set frequency thresholds, apply Apriori operators, and automatically generate association rules without coding. It also offers visualization tools like support-confidence graphs and rule lists for better interpretation. This makes RapidMiner an excellent choice for students, researchers, and businesses looking to extract actionable knowledge from large datasets seamlessly.
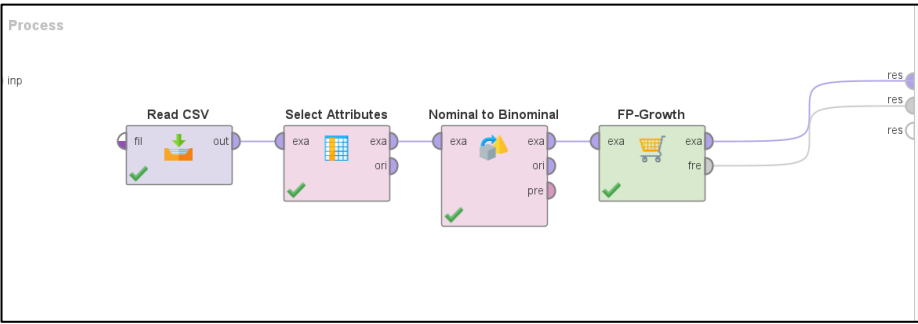
Rapid Miner Design Implementation:

- Import the dataset using the **Read CSV** operator and configure the source file and delimiter settings.

- Use the **Select Attributes** operator to include only categorical attributes relevant for the association analysis, such as ProductSKU or ProductCategory, and <u>exclude numerical columns</u>.

- Apply the **Nominal to Binominal** operator to transform selected nominal (categorical) attributes into a transaction format suitable for association rule mining.

- Add the **FP-Growth** operator to mine frequent itemsets and generate association rules from the binomial (basket-style) data, set ***min_support to 0.5***.

- Connect the output of FP-Growth to the results port to view, analyze, and interpret discovered association rules for further decision-making and business insights.

Inference:

- The highest support is for "Product_SKU = ECLIPSE-EYE-10" and "Product_Category = Serum", indicating these are the most commonly purchased items in the dataset.

- Frequent single items such as "STELLAR-FOUND-03", "AURORA-LIP-01", and "Highlighter" also appear prominently, suggesting these products and categories have strong individual sales footprints.

- The analysis primarily reveals top-selling products and categories, providing actionable insight for promotions, inventory planning, and targeted marketing towards these frequently occurring items.

Results:

No. of Sets: 7	Size	Support	Item 1
Total Max. Size: 1	1	0.110	Product_SKU = ECLIPSE-EYE-10
	1	0.109	Product_Category = Serum
Min. Size: 1	1	0.109	Product_SKU = STELLAR-FOUND-03
Max. Size: 1	1	0.107	Product_SKU = AURORA-LIP-01
Contains Item:	1	0.107	Product_Category = Highlighter
	1	0.106	Product_SKU = ORION-CONCEAL-09
[Update View]	1	0.104	Product_Category = Blush

Conclusion: The experiment successfully demonstrates how to apply association rule mining using RapidMiner and the FP-Growth algorithm. The process involved data import, attribute selection, nominal conversion, and mining frequent itemsets to generate meaningful rules. The output highlights the most frequent products and categories, providing valuable insights for marketing and sales strategies based on product associations. This approach validates the effectiveness of RapidMiner for uncovering hidden purchasing patterns in large transactional datasets.