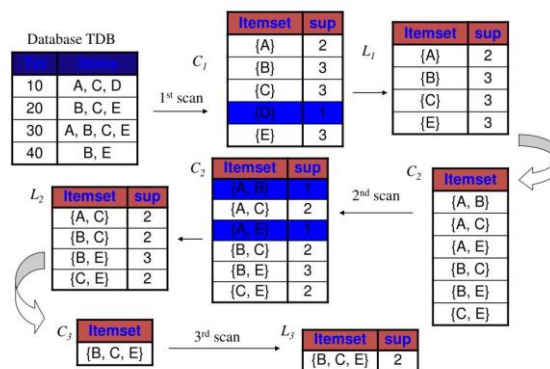


Innovation

Advanced Association Analysis Techniques:

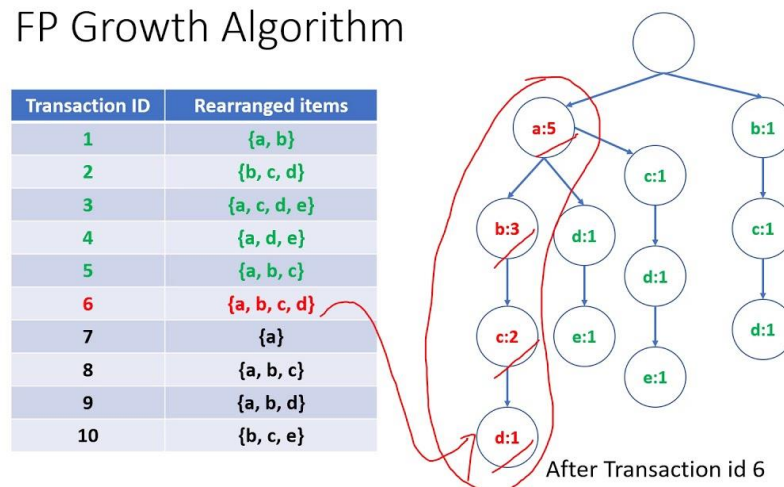
1.A priori Algorithm: A priori is a classic algorithm for discovering frequent item sets in transactional databases. It can be used for market basket analysis to identify associations between products frequently purchased together.

The Apriori Algorithm: An Example



2.FP-Growth Algorithm: FP-Growth is an efficient alternative to A priori for frequent pattern mining. It constructs a compact data structure called an FP-tree to find frequent item sets.

FP Growth Algorithm



3.Sequential Pattern Mining: Instead of just looking at individual item associations, sequential pattern mining identifies patterns in sequences of items. This is useful in analysing sequences like clickstreams, purchase histories, and more.

Sequences in market basket analysis

CID	TID	Transactions
1	100	a,b,c,d
3	111	a,f,d,e
1	122	d,e,p
3	133	b,f,s,a
1	144	b,c,d,e
3	155	a,f,d,e
1	166	a,e,p

We can represent Customer 1 all transactions in a transaction list called Sequence i.e

$$S = \{(a,b,c,d),(d,e,p),(b,c,d,e),(a,e,p)\}$$

4.Association Rule Mining with Constraints: You can add constraints to association rules, such as minimum support or confidence thresholds, to make the discovered patterns more meaningful and actionable.

Association Rule Mining

- Given a set of transactions, find rules that will predict the occurrence of an item based on the occurrences of other items in the transaction

Market-Basket transactions

TID	Items
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

Example of Association Rules

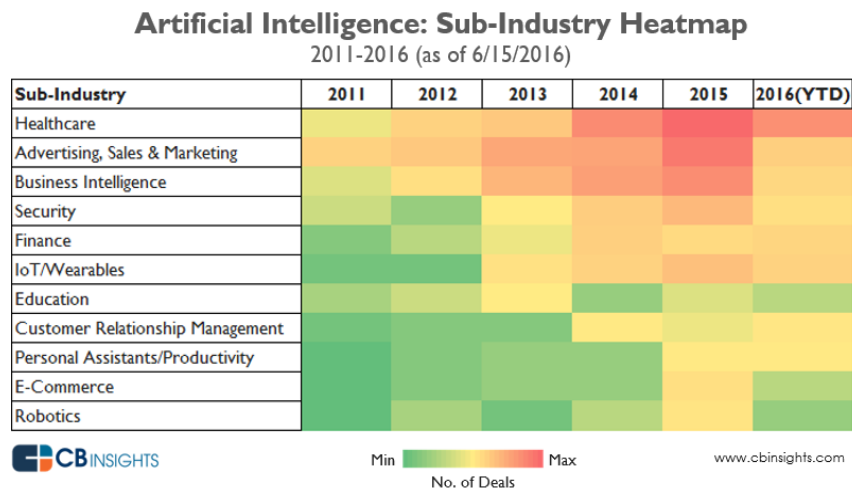
$\{\text{Diaper}\} \rightarrow \{\text{Beer}\},$
 $\{\text{Milk, Bread}\} \rightarrow \{\text{Eggs,Coke}\},$
 $\{\text{Beer, Bread}\} \rightarrow \{\text{Milk}\},$

Implication means co-occurrence, not causality!

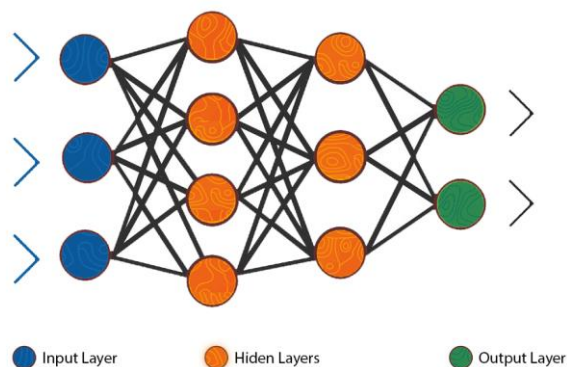
Innovation

Visualization Tools:

1. Heatmaps: Heatmaps can be used to visualize the strength and direction of associations between variables. You can use colour coding to represent the degree of association, making it easy to spot patterns.

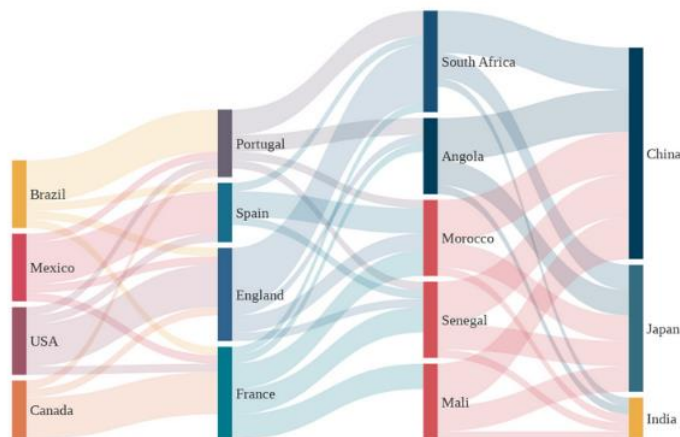


2. Network Graphs: Network graphs are useful for visualizing complex relationships. You can represent associations as nodes and use edges to connect related items. This is helpful for understanding interconnectedness in data.



3.Sankey Diagrams: Sankey diagrams are great for showing flow and transitions between categories or items. They are often used to visualize customer journeys, supply chain flows, and more.

Sankey diagram



4. Scatter Plots and Bubble Charts: These can help visualize associations and correlations between two or more numeric variables. You can use different markers or colors to highlight different categories or attributes.

Scatter Bubble Chart



> **Product 1**
This slide is 100% editable. Adapt it to your needs and capture your audience's attention.

> **Product 2**
This slide is 100% editable. Adapt it to your needs and capture your audience's attention.

> **Product 3**
This slide is 100% editable. Adapt it to your needs and capture your audience's attention.

This graph/chart is linked to excel, and changes automatically based on data. Just left click on it and select "Edit Data".



Innovation

5.Interactive Dashboards: Tools like Tableau, Power BI, and QlikView allow you to create interactive dashboards that provide a dynamic way to explore and present associations in your data. Users can filter, drill down, and interact with the data.

