Problem Statement

**Market Basket Analysis**

There are many data analysis tools available to the python analyst and it can be challenging to know which ones to use in a particular situation. A useful (but somewhat overlooked) technique is called association analysis which attempts to find common patterns of items in large data sets. One specific application is often called market basket analysis. The most commonly cited example of market basket analysis is the so-called “beer and diapers” case. The basic story is that a large retailer was able to mine their transaction data and find an unexpected purchase pattern of individuals that were buying beer and baby diapers at the same time.

Association Analysis : it is an unsupervised learning tool that looks for hidden patterns so there is limited need for data prep and feature engineering.

Solution to Association Analysis

**Support** is the relative frequency that the rules show up. In many instances, you may want to look for high support in order to make sure it is a useful relationship. However, there may be instances where a low support is useful if you are trying to find “hidden” relationships.

**Confidence** is a measure of the reliability of the rule. A confidence of .5 in the above example would mean that in 50% of the cases where Diaper and Gum were purchased, the purchase also included Beer and Chips. For product recommendation, a 50% confidence may be perfectly acceptable but in a medical situation, this level may not be high enough.

**Lift** is the ratio of the observed support to that expected if the two rules were independ. The basic rule of thumb is that a lift value close to 1 means the rules were completely independent. Lift values > 1 are generally more “interesting” and could be indicative of a useful rule pattern.

Data

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **InvoiceNo** | **StockCode** | **Description** | **Quantity** | **InvoiceDate** | **UnitPrice** | **CustomerID** | **Country** |

**Steps Involved**

1. Data Gathering
2. Data Cleaning (Filling Null Values, removing Garbage Data )
3. Feature Training
4. Research Of Market : Models and Algorithms ( With Efficiency)
5. Inferences – Checking Accuracy
6. Results
7. Improve Results – (Comparing Other Models on the basis of Output)

|  |  |  |
| --- | --- | --- |
| **Steps** | **Requirements** | **Action** |
| 1) Data Gathering | Relevent Data Sets, having sales details ,also require test data | have got 2013 BigMart datahaving sales details |
|  |  |  |
| 2) Data Cleaning (Filling Null Values, removing Garbage Data ) | Should have configured ML enabled with Python. | Downloaded Anaconda 3.x ( https://anaconda.org/) |
|  |  |  |
| 3) Feature Training | Should know Algorithms of Supervised and Unsupervised Learning. | Used DecisionTreeRegression- Scikit Learn |
|  |  |  |
| 4) Research Of Market : Models and Algorithms ( With Efficiency) | Should have knowledge of ML models | Classification model |
|  |  |  |
| 5) Inferences – Checking Accuracy | Should Know functions of python | Used Libraries (Pandas, matplotlib,Sklearn,numpy ) for using functions. |
|  |  |  |
| 6) Results |  | got an output of sales prediction of new product with 3 attributes (item type,item id,outlet sales) |
|  |  |  |
| 7) Improve Results – (Comparing Other Models on the basis of Output) |  |  |