**HYPOTHESIS**

Our research work will be implementing with apache spark and scala, and Data can be collected from the following sources: (i) The Electronic News System (ii) The customer service hotline (iii) Word documents (iv) E-commerce transactions and supermarket dataset, The following is a description of these types of data: i. Electronic News System – Internet marketing content which includes featured reports and industry analysis. ii. Customer Service Hotline (Customer Service Data) – The customer service hotline is established to provide appropriate and timely responses to customer demands and complaints. iii. Word Documents – Textual Information. Iv. E-commerce transactions and supermarket Payment transactions of items. From above data sets we will find frequent itemsets and as well association rules using apache spark and scala.

**Expected Outcome:**

Practical outcome can be used to identify items which frequent procure and highest revenues Better understanding through can lead to better marketing plans. In Frequent Itemset mining, we need to identify the following parameters:

Which itemset benefit does the customer seek?

What factors influence demand for the itemset?

What are the important criteria for customers’ choosing this particular item

How often item can buy?

How two items are correlated?

Which itemset can customer select?