## CS6423: Practice Question

Example No.	Color	Туре	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Consider the task of classifying cars as being stolen or not based on the following data:

We want to classify a yellow domestic SUV as being stolen or not. Show how to do this using MapReduce.

## Solution:

Here is the centralised Naïve Bayes dependency graph.



The structure with probability tables is:



The MapReduce architecture for the learning phase is shown below:



Here, we assume that the class node has different code to that of the Attribute nodes, i.e., different counts/probabilities are computed.

The count value that is emitted at the MAP phase consists of the count for each class-value for the Class node, and the count for each (class-value, attribute-value) for Attribute nodes.

The count value that is emitted at the REDUCE phase consists of the probability for each class-value for the Class node, and the probability for each (class-value, attribute-value) for Attribute nodes.