**Assignment**

**A) Context**

* There are 100mn households(HH) and 30 TV channels
* An HH can be tuned in to the TV or not. If tuned in, it can be watching any channel out of the 30 channels

**B) Problem Statement**

Device a M/L algorithm which will accurately predicts the probability of a HouseHold watching a particular channel in during a 30 min dayparts( 48 dayparts for a day)

**C) Training Data**

To train our algorithm, every day we can get the information about a HH about 10 times pinging in a day

Each ping response is a yes or no for whether they are watching a specific channel

**Example**: Is HH 5 tuned in to CNN at 12:43? --- > Response is either "YES" or "No"

**Sample File**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Channel** | **Time** | **Tuned In** |
| **HH 5** | FOX | 11 43 AM | YES |
| **HH 9** | CNN | 3 56 PM | NO |
| **HH 15** | USA | 8 21 PM | YES |
| **HH 33** | CNN | 9 10 PM | NO |

You will get 1000mn rows of data everyday. So eventually the algorithm will have a good amount of training data in few weeks.

So the M/L algo, will keep on using the incremental training data that we get everyday to fine tune the prediction.

\*The time for pinging the status of HH is can fall in any 10 dayparts of the total 48 day parts

\*The number of rows in the one file like this that we will have in one day will be around 1000mn.

**Instructions**

Please give a detailed description of the Model and algorithm that you will use

Specify all the formulae’s that are being used at each step to update values

Make use of diagrams and analytical tools as and when required

Feel free to create your own sample data sets along with others (Cross Validation, Test Data Set) [Volume of actual training data is mentioned above]