# **Customer Profiling using Machine Learning Algorithms**

### Algorithm Used:

- 1) Logistic Regression
- 2) Decision Tree Classifier
- 3) K-means Clustering

#### **Feature Used:**

- a) Age
- b) Sex
- c) Level of Profession
- d) Education

**Target Variable**: Level of Influence (Three classes: 1,2,3)

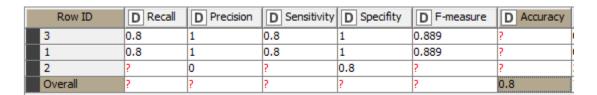
## **Feature Engineering:**

- 1) all the features used are categorical
- 2) Although Type of Location is a categorical feature but it's not useful to build a model due to dominant behavior of 1 (Dhaka)
- 3) Other columns such as Name, Profile are just random string and not useful as a feature at all.

## **Result Analysis:**

## 1) Logistic Regression

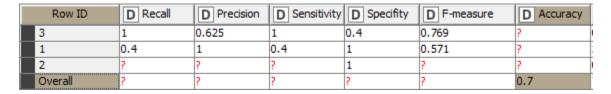
Accuracy achieved 80% with train: test =80:20 split



Row ID	Age	Sex	Level o	Education	S Level of Influence	S Prediction (Level of Influence)
Row3	2	1	2	3	1	1
Row10	3	0	3	3	3	3
Row24	2	0	0	3	1	1
Row29	3	0	3	3	3	3
Row32	3	0	3	3	2	3
Row34	2	1	0	3	1	1
Row35	2	0	1	3	1	1
Row42	3	0	3	3	3	3
Row44	3	0	3	3	3	3
Row49	3	0	2	3	2	1

### 2) Decision Tree Classifier

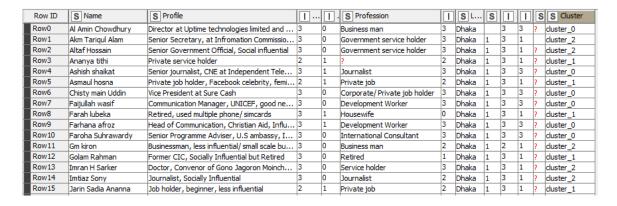
Accuracy achieved 70% with train: test =80:20 split



Row ID	S Age	S Sex	S Level o	S Education	S Level of Influence	S Prediction (Level of Influence)
Row1	3	0	3	3	1	3
Row2	3	0	3	3	1	3
Row4	3	1	3	3	3	3
Row7	3	0	3	3	3	3
Row13	3	0	3	3	1	3
Row15	2	1	2	3	1	1
Row23	3	0	3	3	3	3
Row29	3	0	3	3	3	3
Row33	3	1	3	3	3	3
Row46	2	0	1	3	1	1

## 3) K-means Clustering

Sample Output given below, full excel sheet available separately.



Programming / Simulation Environment: KNIME

By, Md Mahmud Ferdous