

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	D Recall	D Precision	D Sensitivity	D Specifity	D F-measure	D Accuracy
3	0.8	1	0.8	1	0.889	?
1	0.8	1	0.8	1	0.889	?
2	?	0	?	0.8	?	?
Overall	?	?	?	?	?	0.8

Row ID	I Age	I Sex	I Level o...	I 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	D Recall	D Precision	D Sensitivity	D Specifity	D F-measure	D Accuracy
3	1	0.625	1	0.4	0.769	?
1	0.4	1	0.4	1	0.571	?
2	?	?	?	1	?	?
Overall	?	?	?	?	?	0.7

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.

Row ID	S Name	S Profile	I ...	I .	S Profession	I	S L...	S	I	I	S	S Cluster
Row0	Al Amin Chowdhury	Director at Uptime technologies limited and ...	3	0	Business man	3	Dhaka		3	3	?	cluster_0
Row1	Akm Tariqul Alam	Senior Secretary, at Infomation Commissio...	3	0	Government service holder	3	Dhaka	1	3	1		cluster_2
Row2	Altaf Hossain	Senior Government Official, Social influential	3	0	Government service holder	3	Dhaka	1	3	1	?	cluster_2
Row3	Ananya tithi	Private service holder	2	1	?	2	Dhaka	1	3	1	?	cluster_1
Row4	Ashish shaihat	Senior journalist, CNE at Independent Tele...	3	1	Journalist	3	Dhaka	1	3	3	?	cluster_0
Row5	Asmaul hosna	Private job holder, Facebook celebrity, femi...	2	1	Private job	2	Dhaka	1	3	1	?	cluster_1
Row6	Chisty main Uddin	Vice President at Sure Cash	3	0	Corporate/ Private job holder	3	Dhaka	1	3	3	?	cluster_0
Row7	Faijullah wasif	Communication Manager, UNICEF, good ne...	3	0	Development Worker	3	Dhaka	1	3	3	?	cluster_0
Row8	Farah lubeka	Retired, used multiple phone/ simcards	3	1	Housewife	0	Dhaka	1	3	1	?	cluster_1
Row9	Farhana afroz	Head of Communication, Christian Aid, Infl...	3	1	Development Worker	3	Dhaka	1	3	3	?	cluster_0
Row10	Faroha Suhrawardy	Senior Programme Adviser, U.S embassy, I...	3	0	International Consultant	3	Dhaka	1	3	3	?	cluster_0
Row11	Gm kiron	Businessman, less influential/ small scale bu...	3	0	Business man	2	Dhaka	1	2	1	?	cluster_2
Row12	Golam Rahman	Former CIC, Socially Influential but Retired	3	0	Retired	1	Dhaka	1	3	1	?	cluster_1
Row13	Imran H Sarker	Doctor, Convenor of Gono Jagoron Moinch...	3	0	Service holder	3	Dhaka	1	3	1	?	cluster_2
Row14	Imtiaz Sony	Journalist, Socially Influential	3	0	Journalist	2	Dhaka	1	3	1	?	cluster_2
Row15	Jarin Sadia Ananna	Job holder, beginner, less influential	2	1	Private job	2	Dhaka	1	3	1	?	cluster_1

Programming / Simulation Environment: **KNIME**

By,

Md Mahmud Ferdous