

MAPPING THE PATH OF CONVERSION

BY

TUHIN CHATTOPADHYAY, PH.D.

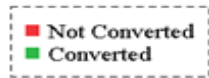
Sl. No.	Dependent Variable	Scale	Categories
1.	Conversion	Nominal	Converted & Not Converted

Sl. No.	Independent Variables	Scale	Categories
1.	Affectv.co.uk Prospecting Clicks	Nominal	Yes & No
2.	Affectv.co.uk Prospecting Impression	Nominal	Yes & No
3.	DART Search. Google Art Generic Clicks	Nominal	Yes & No
4.	DART Search. Google Artist Clicks	Nominal	Yes & No
5.	DART Search. Google Exhibition Brand Clicks	Nominal	Yes & No
6.	exponential.com Prospecting Clicks	Nominal	Yes & No
7.	exponential.com Prospecting Impression	Nominal	Yes & No
8.	guardian.co.uk2 Prospecting Clicks	Nominal	Yes & No
9.	guardian.co.uk2 Prospecting Impression	Nominal	Yes & No
10.	Time Out London Prospecting Clicks	Nominal	Yes & No
11.	Time Out London Prospecting Impression	Nominal	Yes & No
12.	viewlondon.co.uk Prospecting Clicks	Nominal	Yes & No
13.	viewlondon.co.uk Prospecting Impression	Nominal	Yes & No

STATISTICAL TESTS USED:

1. Decision Tree: To delineate the path of conversion
2. Neural Network – Multilayer Perceptron (MLP) and Radial Basis Function (RBF):
To assess the importance of each of the variables.
3. Logistic Regression: To identify the odds of conversion for the different variables.

Conversion



Node 0		
Category	%	n
Not Converted	99.6	2674276
Converted	0.4	11472
Total	100.0	2685748

DART Search Google Artist Clicks
Adj. P-value=0.000, Chi-square=630135.509, df=1

Yes

Node 1		
Category	%	n
Not Converted	64.7	14142
Converted	35.3	7716
Total	0.8	21858

exponential.com Prospecting Impression
Adj. P-value=0.003, Chi-square=9.088, df=1

No

Node 2		
Category	%	n
Not Converted	99.9	2660134
Converted	0.1	3756
Total	99.2	2663890

DART Search Google Exhibition Brand Clicks
Adj. P-value=0.000, Chi-square=20878.125, df=1

No

Node 3		
Category	%	n
Not Converted	64.6	13930
Converted	35.4	7638
Total	0.8	21568

guardian.co.uk2 Prospecting Impression
Adj. P-value=0.021, Chi-square=5.343, df=1

Yes

Node 4		
Category	%	n
Not Converted	73.1	212
Converted	26.9	78
Total	0.0	290

No

Node 5		
Category	%	n
Not Converted	99.9	2659897
Converted	0.1	3656
Total	99.2	2663553

Time Out London Prospecting Clicks
Adj. P-value=0.000, Chi-square=18111.537, df=1

Yes

Node 6		
Category	%	n
Not Converted	70.3	237
Converted	29.7	100
Total	0.0	337

No

Node 7		
Category	%	n
Not Converted	64.5	13860
Converted	35.5	7616
Total	0.8	21476

Yes

Node 8		
Category	%	n
Not Converted	76.1	70
Converted	23.9	22
Total	0.0	92

No

Node 9		
Category	%	n
Not Converted	99.9	2659020
Converted	0.1	3494
Total	99.1	2662514

Yes

Node 10		
Category	%	n
Not Converted	84.4	877
Converted	15.6	162
Total	0.0	1039

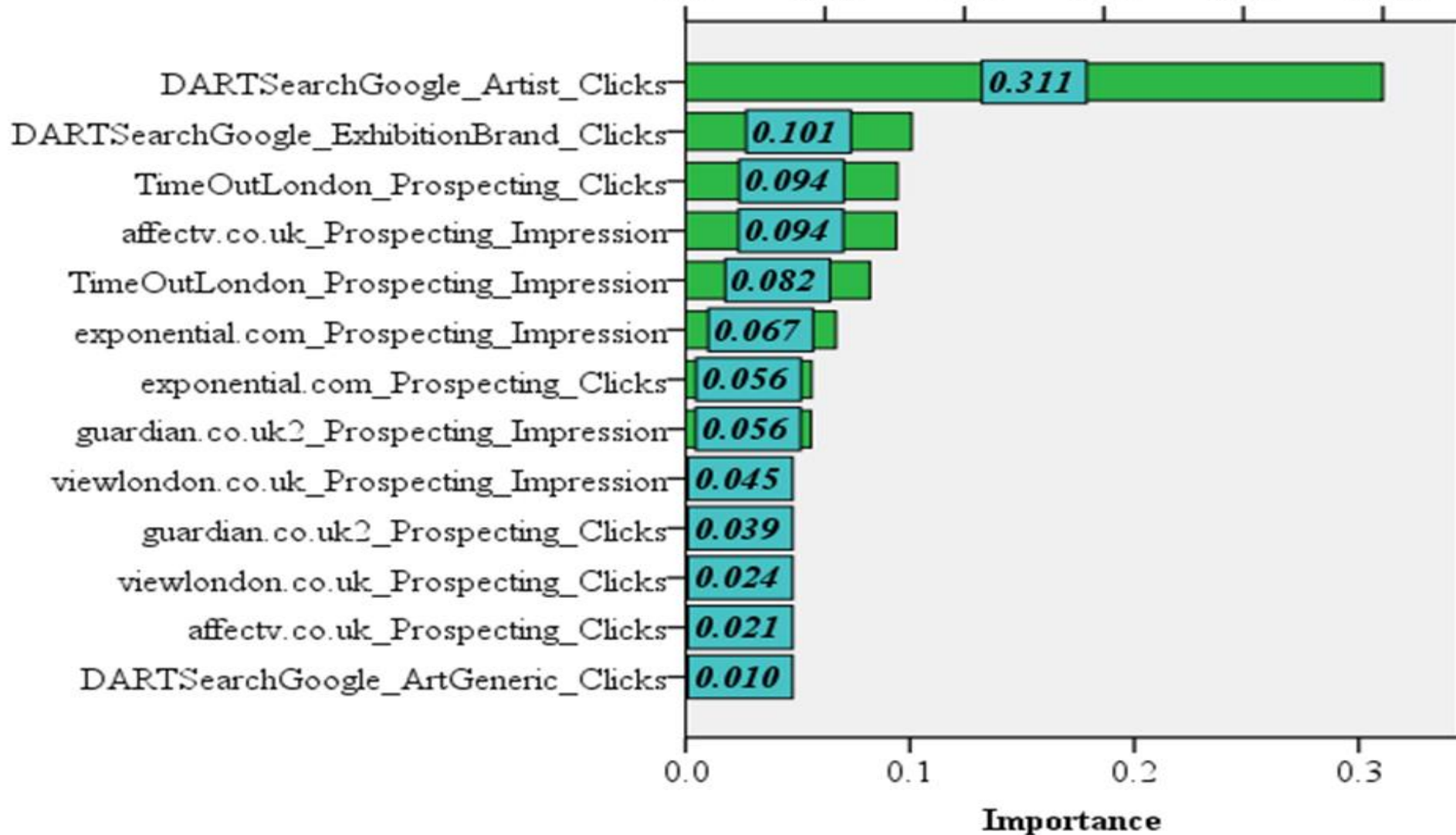
INSIGHTS FROM DECISION TREE

1. The given population is 26,85,748. Out of which 26,74,276 (99.6%) are not converted and only 11,472 (.4%) are converted.
2. The click on DART Search Google Artist is the best predictor of Conversion.
3. Among the persons who clicked on DART Search Google Artist, 7716 (35.3%) persons got converted.
4. Among the persons who didn't click on DART Search Google Artist, the click on DART Search Google Exhibition Brand is the best predictor. The click on DART Search Google Exhibition Brand generates 29.7% conversion.
5. Among the persons who didn't click on DART Search Google Exhibition Brand, the click on Time Out London (15.6%) is the best predictor.
6. Among the persons who clicked on DART Search Google Artist, impression on exponential.com is the best predictor and 26.9% is the conversion rate among the persons who had impression and 35.4% is the conversion among the persons who didn't have impression at exponential.com.
7. Among the persons who didn't have an impression at exponential.com, the impression at guardian.co.uk2 is the best predictor and 23.9% converted from the persons who had the impression and 35.5% got converted who didn't have the impression.

NEURAL NETWORKING - MULTI LAYER PERCEPTRON (MLP)

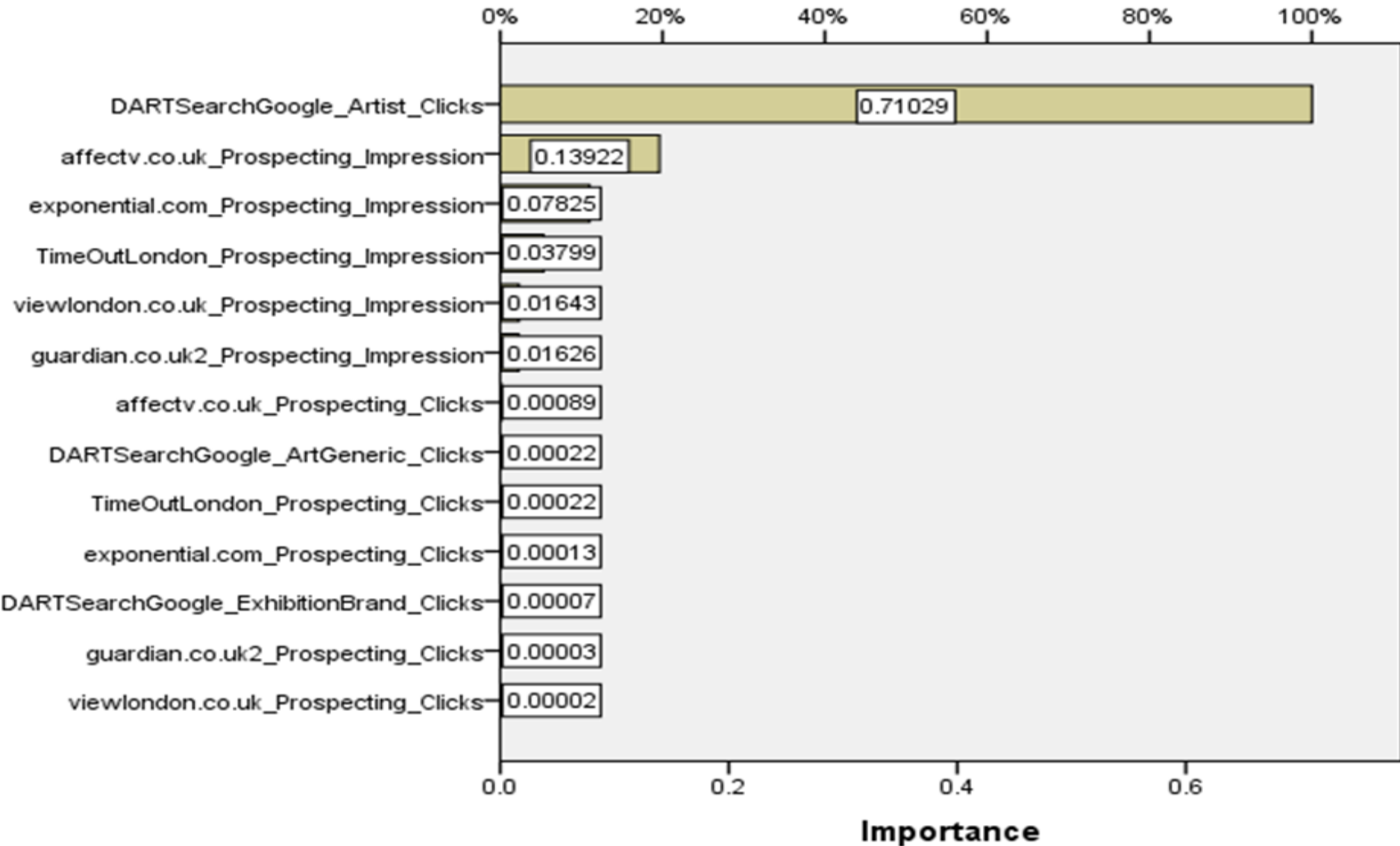
Normalized Importance

0% 20% 40% 60% 80% 100%



NEURAL NETWORKING – RADIAL BASIS FUNCTION (RBF)

Normalized Importance



LOGISTIC REGRESSION

Variables in the Equation						
	B	S.E.	Wald	df	Sig.	Exp(B)
affectv.co.uk Prospecting Clicks	-2.186	.320	46.714	1	.000	.112
affectv.co.uk Prospecting Impression	2.482	.062	1602.497	1	.000	11.970
DART Search Google Art Generic Clicks	-2.953	.131	504.381	1	.000	.052
DART Search Google Artist Clicks	-5.041	.043	13600.361	1	.000	.006
DART Search Google Exhibition Brand Clicks	-4.692	.129	1320.116	1	.000	.009
exponential.com Prospecting Clicks	-1.922	.581	10.958	1	.001	.146
exponential.com Prospecting Impression	1.606	.057	784.157	1	.000	4.981
guardian.co.uk2 Prospecting Clicks	-1.588	.643	6.110	1	.013	.204
Time Out London Prospecting Clicks	-2.973	.089	1114.406	1	.000	.051
Time Out London Prospecting Impression	-1.047	.044	567.519	1	.000	.351
viewlondon.co.uk Prospecting Clicks	-4.130	.266	240.951	1	.000	.016
Constant	16.774	1.017	272.210	1	.000	19274429.100

Variables not in the Equation				
		Score	df	Sig.
Step 2	viewlondon.co.uk Prospecting Impression	.678	1	.410
Step 3	guardian.co.uk2 Prospecting Impression	1.836	1	.175


Critical Insights:

- The odds of conversion are 12 times higher for an impression on affectv.co.uk than they are for subjects who have not had the impression at affectv.co.uk, all other things being equal.
- The odds of conversion are 5 times higher for an impression on exponential.com than they are for subjects who have not had the impression at exponential.com, all other things being equal.

TO CUT THE LONG STORY SHORT...

	Decision Tree	NN - MLP	NN - RBF	LR
DART Search Google Artist Clicks	Highest	Highest	Highest	X
DART Search Exhibition Brand Clicks	High	High	X	X
Affectv.co.uk Prospecting Impression	X	High	High	Highest
exponential.com Prospecting Impression	High	Medium	High	High

SCOPE FOR IMPROVEMENT

1. Generate decision trees, neural network and logistic regression with the click through rates for each of the sites.
 2. Given the revenue of each site, we may generate misclassification cost, gain chart and lift chart.
 3. We may generate sub-models for each site keeping click as the dependent variable and the impression as the independent variable.
 4. We may consider the time-stamp into our models and generate weight based on time before conversion.
- 

THANK YOU

