CAPSTONE PROJECT

Website Behavior Analysis

Submitted by :Deepti Bhanot

Problem Statement

The data shared by client contains

Log data of their User in :
   Problem2_Site_BrowingBehavior.csv

	Timestamp	UserID	Website_section_visited
0	2017-07-26 00:01:40.965	0	product
1	2017-07-26 00:03:18.448	0	product
2	2017-07-26 00:36:59.028	0	default
3	2017-07-26 00:41:17.273	0	product-listing-category
4	2017-07-26 00:45:39.197	0	content

Final Conversion data in: Problem2_FInalConversions.csv

	Timestamp	UserID	Products_Purchased	Cart_Value
0	2017-07-26 00:00:10.465	0	A293532	52.14
1	2017-07-26 00:00:12.301	0	H209597	31.50
2	2017-07-26 00:00:12.388	0	H211370	30.48
3	2017-07-26 00:00:14.389	0	A282331	51.00
4	2017-07-26 00:00:16.837	0	H211410	16.74

We need to perform feature Engineering on given data set to get final dataset

Total_Cart_value	Total_Product_purchased	No_of_times_visited	Time_Spent	BuyProbability	Target
1.000000	1.00000	1.000000e+00	1.000000	0.002161	1.000000
0.000049	0.00000	1.420578e-06	0.839766	0.055338	0.001272
0.000059	0.00000	7.102888e-07	0.484044	0.083122	0.002279
0.000062	0.00000	1.420578e-06	0.434959	0.055338	0.001593
0.000055	0.00005	8.523466e-06	0.697553	0.025416	0.000659

After performing Exploratory Data Analysis on final data set I come to know:
- That feature No of times Visited Contains 155 blank values and % of null values in feature No of times Visited is 38%

- The feature is imputed with value 1

- Target feature is +ve correlated with Total cart value, Total product purchased and No of times visited.

- The target feature is continuous in nature so the problem is of regression.

Model Selection

Models used for analysis are:

•Linear Regression

•Decision Tree Regressor

•KNeighborsRegressor

•SVR

Performance of Models

	MSE	RMSE	MAE	R2
Decision tree	0.005924	0.076967	0.005807	0.936186
LinerRegression	0.000001	0.001067	0.000514	0.660125
KNN	0.584549	0.764558	0.165794	-0.203744
SVM	0.101548	0.318666	0.120319	-261448.869834

The value of MSE,RMSE and MAE is minimum in Linear Regression so the Final Model is:

Target=-0.000811+Total_Cart_value11.290344+Total_Product_purchased*0.269231+No_of_times_visited-10.558645+Time_Spent-0.000114+BuyProbability0.041629

The final audience for the above work is business team s who want to send offers to their website users so that they can be converted and buy products from site .