

Analytics Vidhya Job-a-thon June 2022

APPROACH

Null Value Imputation : Null values were present in products_purchased, signup_date

```
id                0
created_at        0
campaign_var_1    0
campaign_var_2    0
products_purchased 20911
signup_date       15113
user_activity_var_1 0
user_activity_var_2 0
user_activity_var_3 0
user_activity_var_4 0
user_activity_var_5 0
user_activity_var_6 0
user_activity_var_7 0
user_activity_var_8 0
user_activity_var_9 0
user_activity_var_10 0
user_activity_var_11 0
user_activity_var_12 0
buy              0
dtype: int64
```

Upon checking value counts of products_purchased, realised that null values for this column could indicate that no products were purchased by these leads. Hence imputed these null values with zero.

```
2.0    8867
3.0    5024
1.0    3643
4.0     716
Name: products_purchased, dtype: int64
```

Feature Engineering with signup_date : Datatype of signup_date and created_at was string, converted both to datetime and subtracted both to analyse the relationship between the two (if any).

Initial hypothesis was that lead created date will always be after signup date, but after finding the difference between them, identified that hypothesis is not true. As for some cases difference was negative.

r_9	user_activity_var_10	user_activity_var_11	user_activity_var_12	buy	no_of_days
0	0	0	0	0	99 days
0	0	0	0	0	104 days
0	0	0	0	0	-222 days
0	0	0	0	0	1185 days
0	0	1	0	0	207 days
...
0	0	0	0	0	1534 days
0	0	0	0	0	NaT
0	0	0	0	0	NaT
0	0	1	0	0	NaT
0	0	0	0	0	873 days

NaT values indicated that user has not signed up yet, so imputed these with 0 days.

Based on this info classified sign up date as:

1. After : data points where user signed up after being created as lead (potential customer to buy the product)
2. Before : data points user signed up before being created as lead (existing customers who may or may not have bought product before)
3. Not signed up : data points where signed up date is null

This could be a useful feature as conversion% of signed up user is way more than non signed up user.

	buy		
	count	mean	sum
SignedUp			
After	5844.0	0.061431	359.0
Before	18182.0	0.088659	1612.0
Not Signed Up	15135.0	0.001784	27.0

Modelling : Modelling was tried using algorithms like Decision Tree, Random Forest, Gradient Boosting,XGBoost. Used Grid search CV for hyperparameter tuning.