

Leads Case Study

Group Submission:

Suel Ahmed, Ranjeeth, Vasanth and Ujwal

Business Problem

- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.
- Therefore we need to build a model so that the company's CEO can understand the important factors that can be target to convert potential lead to long term clients.

Analysis

Analysis methodology-

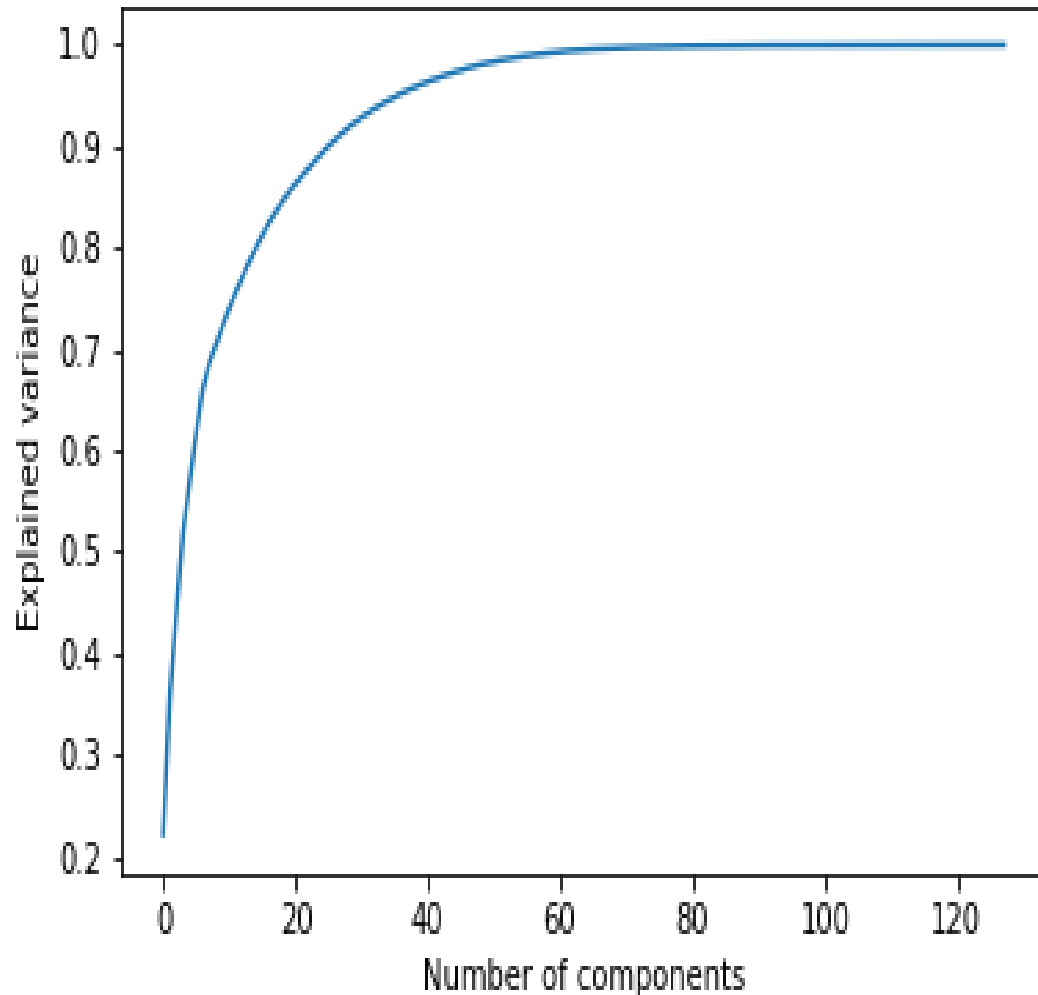
- Read and Clean the data, the threshold for null values was selected to be $> 25\%$
- Remove rows and columns with high null values.
- There are many categorical variables:
 - Variables with two level (Yes/No) map or replace them with 1 and 0 respectively
 - Variables with more than two levels, create dummy variables for analysis
 - After creation of dummy variables, delete the original variablesStandardize the numerical variables using the Standard scaler so that mean=0

Continued

- Standardize the dataset i.e one unique scale for faster analysis.
- Check for correlation , for multi-collinearity.
- Create principal components with an aim to reduce dimensionality
- Check scree plot, to select the number of variable that explain the max variance in dataset.
- Run logistic regression on PC train data set.
- Predict for Test dataset.

Continued

SCREE PLOT



- Metrics(AUC) are calculated to gauge effectiveness of the model in predicting the test dataset
- If $AUC > 0.8$, the model can be assumed fit to predict the lead conversion with a high success
- Also ran a RFE and created an alternate model that can be tested further for specificity , precision and other metrics.

Final Results

The top variables to look out for are:

1. **Lead source Referral**
2. **Lead source welingak website**
3. **Last activity unreachable**