

Lead Conversion Prediction

Summary Report

(Logistic Regression Assignment)

Devesh Singh & Archana Prabhakar

August 2019

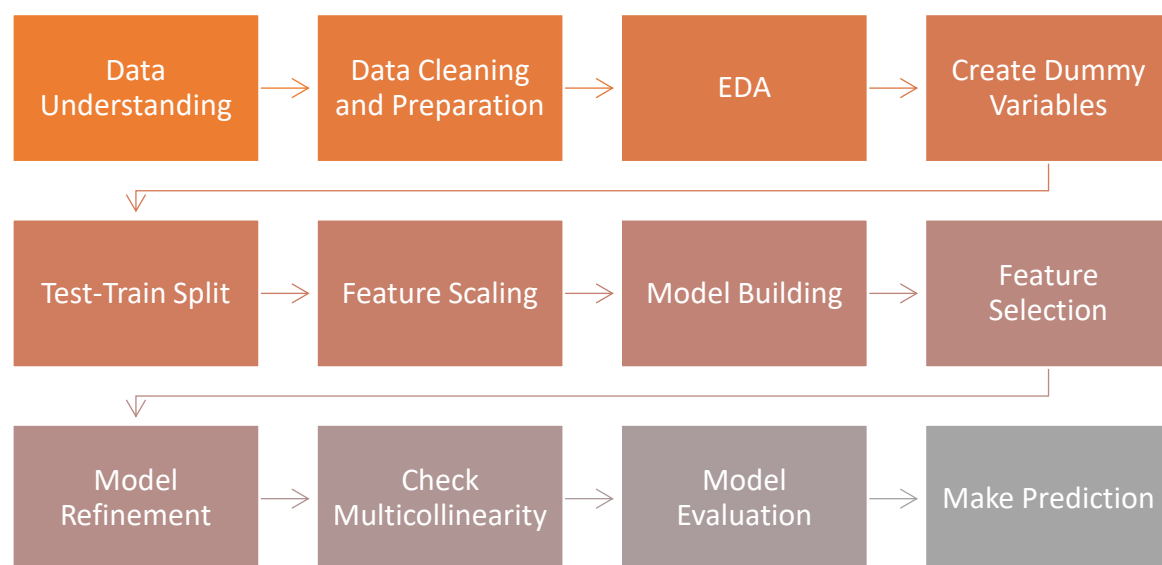
Problem Statement

- X Education, which sells online courses to industry professionals, has a poor Lead conversion rate (30%) and would like to improve it to 80%

Goal

- Build a Logistic Regression Model which can predict which leads are the “Hot” leads (has a high Lead Score)
- Provide recommendations on the what can be done to improve the lead conversion rate

Approach



Some of the important steps are explained in detail below –

Based on the various data quality checks, following actions have been taken -

- Handle missing values
 - Drop Columns where >30% of missing data (E.g. Assymetrique value columns)
 - Impute with mode() where appropriate (E.g. Lead Source, Occupation)
 - Impute with mean() where appropriate (E.g. Total visits)
 - Impute with “Unknown” where appropriate (E.g. City, Specialization)
- Drop columns which provide no additional information/variance to the model building process (E.g. Country, What Matters most)
- Within a column, merge values which have no significant number of rows (E.g. Lead Source values like Social Media, bing etc. merged into Misc_Sources)
- Map binary values to 1 and 0 (E.g. Do Not Call, Search)
- Clean-up of values (E.g. Google and google)

Model Evaluation Metrics

- **Metrics (Train data)**
 - Accuracy – 0.90
 - Sensitivity – 0.89
 - Specificity – 0.90
 - AUC (Based on the RoC curve) – 0.96
 - Precision – 0.85
 - Recall – 0.89
- **Metrics (Test Data)**
 - Accuracy – 0.90
 - Sensitivity – 0.90
 - Specificity – 0.90
- **Top 3 Variables**
 - Lead Source -Welingk_Website
 - Last Activity – SMS Sent, Will revert after reading email
 - Tags – Lost to EINS and Closed by Horizonn

Final Recommendations

Sales team should aggressively reach out to potential candidates via call/email

Ensure that communication is kept on via SMS as well

Since a lot of people seem to indicate that they will revert after reading the email, there should be more aggressive follow-up once the emails are sent

Since the Welingk Website seems to be a huge source of leads, the digital advertising of the same can be increased to ensure more traffic on the site

All leads that are recently updated have more potential for conversion versus the inactive ones, hence the sales team should focus on such leads

X Education also seems to perform well when the Leads are closed by Horizonn, so Horizonn should be more actively engaged in pursuing leads

They also seem to be losing business to EINS, this can be researched to find the potential causes for the loss of business

Key Learnings

1. Data cleaning/feature engineering is a major part of the problem solving
2. EDA provides insights of data not just from a business perspective but also from a raw data perspective
3. Modelling needs to be done iteratively to arrive at the best possible set of predictors
4. Predictors are not always evident – we need to look closer at the data columns and the values in them, since each of the values could drive an action for achieving the goal