LEAD SCORING CASE STUDY

PROBLEM STATEMENT

X Education is an online education provider for industry professionals, selling courses through their website and various marketing channels. The company acquires leads through website forms, videos, referrals, and other sources. The current lead conversion rate is 30%, which is considered low, and the company wants to identify the most promising leads to increase their conversion rate to around 80%.

The objective of the project is to build a model that assigns a lead score to each lead based on their attributes and activities, indicating their likelihood of converting into a paying customer. The model should identify the key factors that influence lead conversion and help the sales team focus their efforts on the most promising leads.

The data provided includes information on various attributes of the leads, such as their demographic information, website activity, and communication history. The model should be able to predict which leads are most likely to convert based on these attributes and activities.

The ultimate goal of the project is to improve the lead conversion rate and increase revenue for X Education by efficiently identifying and nurturing the most promising leads.

GOALS OF CASE STYDY

The main goal of this case study is to build a logistic regression model that can accurately assign a lead score between 0 and 100 to each of the leads. This lead score will indicate the likelihood of a lead converting into a paying customer. A higher score will indicate that the lead is hot and more likely to convert, while a lower score will indicate that the lead is cold and less likely to convert. This model will help the company to identify the most promising leads and prioritize their efforts accordingly, which can ultimately lead to a higher conversion rate and increased revenue.

STRATEGY

- Import data
- Clean and prepare the acquired data for further analysis
- Exploratory data analysis for figuring out most helpful attributes for conversion
- Scaling features
- Prepare the data for modal building
- Build a logistic regression modal
- Assign a lead score for each leads
- Test the model on train set
- Evaluate model by different measures and metrics
- Test the model on test set
- Measure the accuracy of the model and other metrics for evalution

EXPLORATORY DATA ANALYSIS

LEAD SOURCE VS CONVERTED

The lead source is the channel or method through which a lead is generated, such as through a Google search, a referral, or by filling out a form on the company's website. On the other hand, converted refers to whether or not a lead has actually become a paying customer.

DO NOT EMAIL VS CONVERTED

In the context of this case study, "Do not email" refers to leads who have explicitly opted out of receiving emails from X Education. The variable "converted" refers to whether or not a lead has become a paying customer.

LAST ACTIVE VS CONVERTED

SMS has shown to be a promising method for getting higher confirmed leads, emails also has high conversions.

DO NOT CALL VS CONVERTED

Most lead prefer not to informed through phone

LAST NOTABLE ACTIVITY VS CONVERTED

Most leads are with message. Email also induce leads.

A FREE COPY OF MASTERING THE INTERVEIW VS COPNVERTED

Leads prefer less copies of interviews

SPECIALIZATION VS CONVERTED

Most of the lead have no information about specialization.

On the other hand,

Marketing management, human resources management has high conversion rates. People from these specializations can be promising lead

LEAD ORIGEN VS CONVERTED

Landing page submission has had high lead conversion

DIGITAL ADVERTISEMENTS VS CONVERTED

Based On The Above Graph Digital

Advertisements Do Not Have Promising Leads

THROUGH RECOMMENDATIONS VS CONVERTED

From the above graph, recommendations are not a good source for promising leads

SEARCH VS CONVERTED

The above graph shows search are not good source of leads

MAGAZINE VS CONVERTED

Magazines do not have higher conversion rate

TOTAL TIME SPENT ON WEBSITES VS CONVERTED

People spending higher than average time are promising leads

TOTAL VISITS VS CONVERTED

Higher total visits have a slight higher chances of being a promising lead

MODEL BULDING

- Splitting into train and test set
- Scale variables in train set
- Build the first model
- Use RFE to eliminate less relevant variables
- Build the next model
- Eliminate variables based on high p-values
- Check VIF value for all the existing columns
- Predict using train set
- Evaluate accuracy and other metric
- Predict using test set
- Precision and recall analysis on test predictions

MODEL EVOLUTION (TRAIN)

MODEL EVALUATION (TEST)

CONCLUSION

1. We evaluated the model's performance using Sensitivity-Specificity and Precision-Recall metrics to predict the target variable.
2. We selected the optimal cut-off based on Sensitivity and Specificity to calculate the final prediction.
3. The accuracy, Sensitivity, and Specificity values of the test set were approximately 81%, 79%, and 82%, respectively.
4. These values were similar to the respective values obtained from the trained set, suggesting that the model generalizes well to new data.
5. The lead score calculated on the trained set indicated that the conversion rate on the final predicted model is around 80%.
6. Based on these results, we conclude that the model is good at predicting the target variable.
7. However, further validation and testing are necessary to optimize its performance and fully validate its usefulness.
8. It's also important to consider other metrics and trade-offs between them to evaluate the model's performance more comprehensively.
9. Overall, the model has the potential to be useful in practical applications requiring accurate prediction of the target variable.