Lead Score Case Study Summary Report

PRIMARY OBJECTIVE

- 1. The company's primary objective is to identify the most promising leads, referred to as "Hot Leads," and improve the lead conversion process. To achieve this, the company requires a lead scoring model that assigns a lead score to each potential customer based on their likelihood of conversion. Leads with higher lead scores are expected to have a greater chance of converting into paying customers, while those with lower scores are anticipated to have a lower probability of conversion.
- 2. The CEO has provided a rough estimation of the target lead conversion rate, which stands at 80%. This implies that the company aims to achieve an 80% success rate in converting leads into actual customers. By implementing an efficient lead scoring system, X Education aims to prioritize and focus its resources on the most promising leads, ultimately increasing the overall conversion rate and maximizing the return on investment for its marketing and sales efforts.

APPROACH FOLLOWED

1. Data Cleaning and Preparation:

- 1) Acquiring data from the source.
- 2) Preparing and cleansing the data to ensure its suitability for analysis.
- 3) Identifying and eliminating duplicate entries.
- 4) Addressing outliers in the dataset.
- 5) Conducting exploratory data analysis to gain insights and understanding from the data.

2. Data splitting and Feature scaling

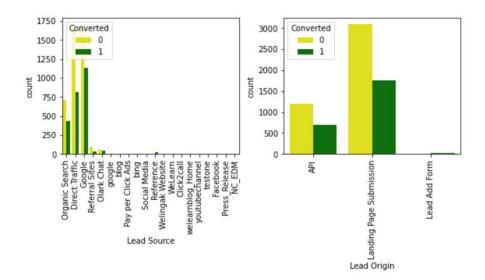
- 1) The process of dividing the data into separate test and train datasets.
- 2) Performing feature scaling on numerical variables to standardize their values.

3. Model Building

- 1) Employing RIF, VIF, and p-value techniques for feature selection.
- 2) Utilizing Logistic Regression to identify the optimal model.
- 3) Computing diverse evaluation metrics to assess model performance.

Result:

- I. Calculate the lead score and verify if the target final prediction achieves a conversion rate higher than 80%.
- II. Assess the performance of the final prediction using the test set.



- 1) The highest number of leads originates from Google and Direct Traffic sources.
- 2) The leads coming from Reference and Welingak Website exhibit the highest conversion rate.
- 3) Although API and Landing Page Submission generate a substantial number of leads, their conversion rate is relatively low, standing at approximately 30%.
- 4) Despite the Lead Add Form contributing a small number of leads, it boasts an impressively high conversion rate.

Model Building

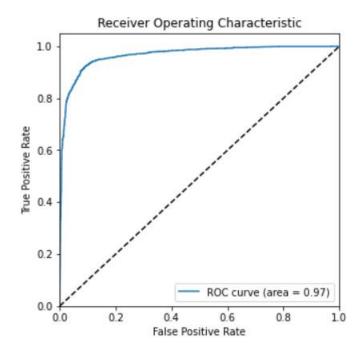
The dataset was divided into training and test sets with a split ratio of 70:30.

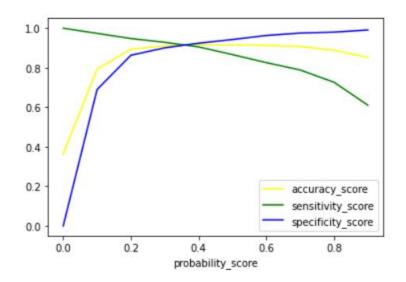
Recursive Feature Elimination (RFE) was utilized to select the most influential features, resulting in the identification of the top 15 variables.

A model was constructed by eliminating variables with a p-value greater than 0.05 and a VIF (Variance Inflation Factor) exceeding 5.

Subsequently, predictions were made on the test dataset.

The overall accuracy achieved by the model was recorded at 91.0%.





Conclusion

The logistic regression model serves to predict the probability of customer conversion. We have evaluated both sensitivity-specificity and precision-recall metrics, but we selected the optimal cutoff based on sensitivity-specificity for the final predictions.

The Lead Score calculated indicates that the conversion rate of the final predicted model is approximately 91% on the test data, slightly lower than the 95% achieved on the train data. Nevertheless, in business terms, this model has the adaptability to meet the company's future requirements.

The top variables contributing to lead conversion in the model are identified as follows:

Tags_Lost to EINS

Tags_Closed by Horizzon

Lead Quality_Worst
Overall, the model demonstrates promising performance and seems to be effective in predicting lead conversions successfully.