Valkal Bairagi
DS C57

Lead Scoring Case Study Summary

The case study delves into the practical challenges faced by X Education, an online education company catering to industry professionals. The company strategically promotes its courses across various online platforms, including popular search engines like Google. Upon landing on the website, potential customers engage with the content by perusing courses, filling out forms, or watching informative videos. A crucial aspect of their marketing strategy involves classifying individuals who provide their contact information as leads. Additionally, the company generates leads through referrals.

Once leads are in the company's database, the sales team initiates contact through phone calls and emails, aiming to convert these leads into paying customers. However, the lead conversion process is intricate, with only a fraction of leads successfully converting. X Education typically experiences a lead conversion rate of approximately 30%. In a bid to enhance this conversion rate, the company sought to identify the most promising leads, referred to as 'Hot Leads.' The goal was to develop a model capable of assigning a lead score to each prospect, predicting a higher likelihood of conversion for leads with higher scores and a lower likelihood for those with lower scores. The CEO set an ambitious target lead conversion rate of around 80%.

The foundation for this initiative was a dataset encompassing about 9000 data points, offering insights into various attributes such as Lead Source, Total Time Spent on Website, Total Visits, and Last Activity. The pivotal target variable was the 'Converted' column, indicating whether a lead from the past had been successfully converted (1) or not (0).

The initial phase involved a meticulous analysis of the dataset to identify relevant variables and determine preprocessing requirements. Notably, it was observed that certain categorical variables contained a level labeled 'Select,' akin to a null value, necessitating careful handling. Furthermore, variables like 'Lead Source' exhibited a multitude of unique categories, prompting the creation of a new category for values with a frequency less than 100 to avoid potential impacts on the model's performance.

Post preprocessing, logistic regression was employed to construct a predictive model for assigning lead scores. The model underwent training on 70% of the data and was subsequently tested on the remaining 30%. Evaluation metrics, including accuracy, precision, recall, and F1-score, were employed to assess the model's performance.

The final model demonstrated an impressive accuracy of 90.92%. This implies that the model accurately identified 73% of potential customers who eventually converted into paying customers. Moreover, the model unearthed critical factors influencing lead conversion, including Total Time Spent on Website, Total Visits, and Lead Origin.

In summary, the case study illustrates the strategic approach undertaken by X Education to optimize lead conversion rates through data-driven insights and predictive modeling. The successful implementation of this approach resulted in a substantial improvement in identifying high-potential leads and ultimately enhancing the company's overall conversion rate.