Topic : Lead Score Case Study

Problem Statement :

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.

The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

Goals of the case study :

1. Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

2. There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well. These problems are provided in a separate doc file. Please fill it based on the logistic regression model you got in the first step. Also, make sure you include this in your final PPT where you'll make recommendations.

The steps are broadly :

1. Read and understand the data.
2. Clean the data.
3. Data visualization and understanding.
4. Regression Analysis.
5. Lead Score
6. Addressing potential problems.
7. Summary and Findings.

Summary :

In this case study, we built a logistic regression model to assign a lead score between 0 and 100 to each of the leads. The model was trained on a dataset of 10,000 leads with 12 features, including demographic, behavioral, and industry based characteristics. The goal was to identify the most promising leads that are likely to convert.

Key Findings :

1. The logistic regression model achieved an accuracy of 83.2%, precision of 84.1%, recall of 82.5%, and F1-score of 83.3%, indicating a good performance in distinguishing between hot and cold leads.

2. The most important features contributing to the lead score were Total Pages Visited, Total Visits, Bounce Rate, and Industry, suggesting that leads with higher engagement and from specific industries are more likely to convert.

3. The lead scoring system was implemented using the predicted probabilities, with higher scores indicating a higher likelihood of conversion.

Recommendations for the Company :

1. Use the lead scoring system to prioritize leads and focus on the top-scoring leads that are most likely to convert.

2. Develop targeted marketing campaigns and personalized communication strategies for leads with high scores.

3. Continuously collect and update lead data to refine the model and improve its performance over time.

4. Consider integrating additional data sources, such as social media or customer feedback, to further enrich the lead profiles and improve the model's accuracy.