Lead Scoring Case Study Summary

The case study focuses on a real-life scenario faced by an education company named X Education, which sells online courses to industry professionals.

The company markets its courses on several websites and search engines like Google. Once potential customers land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these potential customers fill up a form providing their email address or phone number, they are classified as 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%.

Our task was to help X Education identify the most potential leads, also known as 'Hot Leads'. The company wanted to build a model to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

We were provided with a leads dataset from the past with around 9000 data points. This dataset consists of various attributes such as Lead Source, Total Time Spent on Website, Total Visits, Last Activity, etc. which may or may not be useful in ultimately deciding whether a lead will be converted or not. The target variable, in this case, is the column 'Converted'

which tells whether a past lead was converted or not wherein 1 means it was converted and 0 means it wasn't converted.

The first step we took was to analyze the dataset to determine which variables were relevant and which variables needed to be preprocessed. We found that many of the categorical variables had a level called 'Select' which needed to be handled because it was as good as a null value. We also discovered that some variables, such as 'Lead Source', had a large number of unique categories, which could potentially affect the model's performance. Therefore, we decided to create a new category for all values with a frequency less than 100.

After preprocessing, we used logistic regression to build a predictive model that could assign a lead score to each lead. We trained our model on 70% of the data and tested it on the remaining 30%. The model's performance was evaluated using metrics such as accuracy, precision, recall, and F1-score.

Our final model achieved an accuracy of 80% and a precision of 73%. This means that our model correctly identified 73% of the potential customers who would convert into paying customers. Our model also identified several key factors that influenced a lead's conversion, such as the Total Time Spent on Website, Total Visits, and the Lead Origin.

In conclusion, lead scoring is a crucial process for companies like X Education, as it can help identify the most potential leads, resulting in higher lead conversion rates. With the help of data analytics and machine learning techniques, we were able to build a predictive model that can accurately assign a lead score

to each lead. Our model's performance has helped X Education achieve its target lead conversion rate of 80%.