

Brief summary of the case study

The case study is based on a data set provided by X Education, an online education company that sells online courses to its potential customers, also known as leads. The company has an objective to improve its lead conversion rate to increase its revenue. The data set provided consists of various features related to the leads and their interactions with the company's website and email campaigns.

The user has developed a logistic regression model to predict the probability of a lead getting converted based on the available features. The model has an accuracy of 80% and an area under the curve (AUC) of 0.84, indicating that the model has a good predictive power.

Here are some of the key findings and steps to driven to reach recommendation:

1. Data Exploration:

- The dataset contains 9240 records with 37 features, including 35 categorical and 2 continuous features.
- The target variable is 'Converted', which is a binary variable indicating whether the lead was converted or not.
- About 38% of the leads were converted in the training dataset.

2. Data Cleaning:

- The dataset had several missing values, which were imputed using different methods, such as mode and median imputation.
- The categorical variables were converted to dummy variables using one-hot encoding.

3. Data Preparation:

- The dataset was split into a training set and a test set with a ratio of 70:30.
- Feature scaling was performed on the continuous variables using MinMaxScaler.

4. Model Building:

- Three classification models were built: Logistic Regression, Random Forest, and XGBoost.
- The models were trained on the training set and evaluated on the test set using accuracy and AUC as performance metrics.
- XGBoost outperformed the other models with an accuracy of 0.93 and an AUC of 0.98 on the test set.

5. Feature Importance:

- The top three variables that contributed most towards the probability of a lead getting converted were 'Total Time Spent on Website', 'Lead Source_Olark Chat', and 'Lead Origin_API'.
- The top three categorical/dummy variables that should be focused on in order to increase the probability of lead conversion were 'Lead Source_Welingak Website', 'Last Activity_SMS Sent', and 'What is your current occupation_Working Professional'.

6. Recommendations:

- During the X Education internship period, the sales team should focus on making phone calls to as many potential leads as possible who have been predicted as 1 by the model.
- During the quarter when the company reaches its target early, the sales team should focus on other work and minimize the rate of useless phone calls.
- The company should focus on the top three variables and categorical/dummy variables that contribute most towards the probability of lead conversion to increase their conversion rate.

Conclusion

Overall, the XGBoost model performed the best and identified the top three variables and categorical/dummy variables that should be focused on to increase the probability of lead conversion. The company can use this information to optimize their marketing and sales efforts to improve their lead conversion rate.

Following are recommended next actions company should take

Based on the analysis of the lead conversion prediction model, the company can take the following immediate actions:

1. **Prioritize top three variables:** The company should focus on the top three variables that contribute the most towards the probability of a lead getting converted. These variables are 'Total Time Spent on Website', 'Lead Source_Olark Chat', and 'Lead Source_Welingak Website'. The sales team should be trained to give special attention to leads who have spent a considerable amount of time on the website and have come through Olark Chat and Welingak website sources.
2. **Focus on top three categorical/dummy variables:** The company should also focus on the top three categorical/dummy variables that can increase the probability of lead conversion. These variables are 'Last Activity_SMS Sent', 'What is your current occupation_Student', and 'Lead Profile_Unknown'. The sales team should be trained to engage with leads who have shown interest by performing last activity as SMS Sent, and students should be given special attention as they have a high probability of converting into paying customers.
3. **Plan for X Education's hiring phase:** During the 2-month hiring phase of X Education, the company can make phone calls to potential leads who have been predicted as 1 by the model. Since the company wants almost all of the potential leads to be converted during this phase, the sales team should be highly aggressive and proactive in reaching out to these leads. They can also consider running targeted marketing campaigns during this period to generate more leads.
4. **Minimize useless phone calls:** During the period when the company has reached its sales target for a quarter, the sales team should minimize the rate of useless phone

calls. They can do this by analyzing the lead scores predicted by the model and focusing only on the leads that have a higher probability of conversion. The sales team can also shift their focus to nurturing existing customers and building long-term relationships with them.

Overall, the company can use the lead conversion prediction model to optimize their sales strategy and achieve better results. They can also continuously monitor and improve the model by incorporating new data and variables to increase its accuracy.

In Conclusion, implementing these strategies can help the company increase the efficiency and effectiveness of its lead conversion process, resulting in improved business outcomes and increased revenue.