Summary

X Education gets a lot of leads, its lead conversion rate is very poor at around 30%. The company requires us to build a model wherein we need to assign a lead score to each of the leads such that the customers with a higher lead score have higher conversion chance. CEO's target for lead conversion rate is around 80%.

Data Cleaning:

● Columns with >40% nulls were dropped. Value counts within categorical columns were checked to decide appropriate action: if imputation causes skew, then column was dropped, created new category (others), impute high frequency value, drop columns that don't add any value.

● Numerical categorical data were imputed with mode and columns with only one unique response from customer were dropped.

● Other activities like outliers' treatment, fixing invalid data, grouping low frequency values, mapping binary categorical values were carried out.

EDA:

● Data imbalance checked- only 38.5% leads converted.

● Performed univariate and bivariate analysis for categorical and numerical variables. 'Lead Origin', 'Current occupation', 'Lead Source', etc. provide valuable insight on effect on target variable.

● Time spend on website shows positive impact on lead conversion.

Data Preparation:

● Created dummy features (one-hot encoded) for categorical variables

● Splitting Train & Test Sets: 70:30 ratio

● Feature Scaling using Standardization

● Dropped few columns, they were highly correlated with each other

Model Building:

● Used RFE to reduce variables from 48 to 15. This will make dataframe more manageable.

● Manual Feature Reduction process was used to build models by dropping variables with p – value > 0.05.

● Total 3 models were built before reaching final Model 4 which was stable with (p-values < 0.05). No sign of multicollinearity with VIF < 5.

● logm4 was selected as final model with 12 variables, we used it for making prediction on train and test set.

Model Evaluation:

● Confusion matrix was made and cut off point of 0.345 was selected based on accuracy, sensitivity and specificity plot. This cut off gave accuracy, specificity and precision all around 80%. Whereas precision recall view gave less performance metrics around 75%.

● As to solve business problem CEO asked to boost conversion rate to 80%, but metrics dropped when we took precision-recall view. So,

we will choose sensitivity-specificity view for our optimal cut-off for final predictions

● Lead score was assigned to train data using 0.345 as cut off.

Making Predictions on Test Data:

● Making Predictions on Test: Scaling and predicting using final model.

● Evaluation metrics for train & test are very close to around 80%.

● Lead score was assigned.

● Top 3 features are: o Lead Source_Welingak Website o Lead Source_Reference o Current_occupation_Working Professional

Subjective Questions and their Answers

Question 1:- Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?

Answer:- The top three variables based on the final model which contributes most towards the probability of lead getting converted are:

- Lead Source_Welingak Website: 5.39
- Lead Source_Reference: 2.93
- Current_occupation_Working Professional: 2.67

Question 2:- What are the top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion?

Answer: The top 3 categorical/dummy variables in the model that be focused most in order to increase the probability of lead conversion are:

- Lead Source_Welingak Website: 5.39
    - Weshould focus on more budget/spend on Welingak Website in terms of advertising, etc. to attract more leads.
- Lead Source_Reference: 2.93
    - We can provide discounts for providing references that convert to lead to encourage more references.

● Current_occupation_Working Professional: 2.67

○ We should develop tailored messaging and engage working professionals through communication channels based on their engagement impact

Question 3:- X Education has a period of 2 months every year during which they hire some interns. The sales team, in particular, has around 10 interns allotted to them. So during this phase, they wish to make the lead conversion more aggressive. So they want almost all the potential leads (i.e. the customers who have been predicted as 1 by the model) to be converted and hence, want to make phone calls to as much of such people as possible. Suggest a good strategy they should employ at this stage.

Answer:- To make lead conversion more aggressive during the intern-hiring period, X Education can employ the following strategy based on the given variables and their coefficients:

● Focus on leads with high potential: Based on the given coefficients, leads from the following sources have a higher likelihood of conversion:

o Welingak Website: 5.388662

o Reference: 2.925326

o Working Professional: 2.669665

● Thus, the sales team should prioritize calling leads from these sources during the intern-hiring period.

● Leverage effective communication channels: Leads who have been sent SMS messages and have opened the emails are also more likely to convert. The coefficients for Last Activity_SMS Sent and Last Activity_Email Opened are 2.051879 and 0.942099, respectively. Therefore, the sales team should prioritize calling leads who have been sent SMS messages or have opened emails from X Education.

● Maximize website engagement: Total Time Spent on the Website is also a good indicator of the lead's interest in X Education's services, with a coefficient of 1.049789. Therefore, the sales team should also prioritize calling leads who have spent a significant amount of time on the website.

● Maintain a multi-channel approach: Finally, the sales team should also make sure to follow up with leads who have interacted with X Education through multiple channels. For example, leads who have used the Olark Chat feature on the website may not have spent as much time on the website, but may still be interested in X Education's services. Therefore, the sales team should make sure to follow up with leads who have used multiple channels to interact with X Education.

In summary, to make lead conversion more aggressive during the intern-hiring period, X Education should focus on leads from high-potential sources, leverage effective communication channels, maximize website engagement, and maintain a multi-channel approach.

Question 4:- Similarly, at times, the company reaches its target for a quarter before the deadline. During this time, the company wants the sales team to focus on some new work as well. So during this time, the company's aim is to not make phone calls unless it's extremely necessary, i.e.. they want to minimize the rate of useless phone calls. Suggest a strategy they should employ at this stage?

Answer:- To minimize the rate of useless phone calls when the company reaches its sales target for a quarter before the deadline, the sales team can employ the following strategy:

● Focus on lead nurturing activities such as personalized emails, SMS's and targeted newsletters.

● Sending automated SMS to customers that have very good likelihood of getting converted

● Collaborate with the sales team, management, and data scientists to teams to fine-tune the model and gather feedback on what worked and what didn't.

● Make the strategy for providing discounts or incentives to potential customers to encourage them to take action.

● Focus on building relationships with potential customers through other communication channels like email, social media, or chatbots.
● Gather feedback from existing customers to improve the quality of the leads generated and optimize the conversion rate.