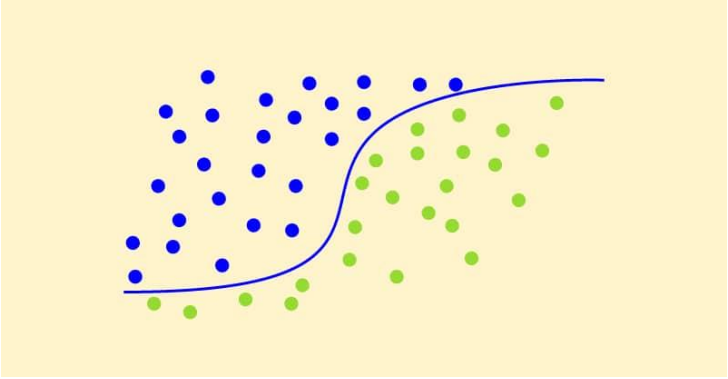
LEAD SCORING CASE STUDY

Based on Logistic Regression



Summary Report

BY

Monalisa Banerjee

Swetha Mennde

Amit kumar

|

Student of IIIT - Bangalore and UpGrad

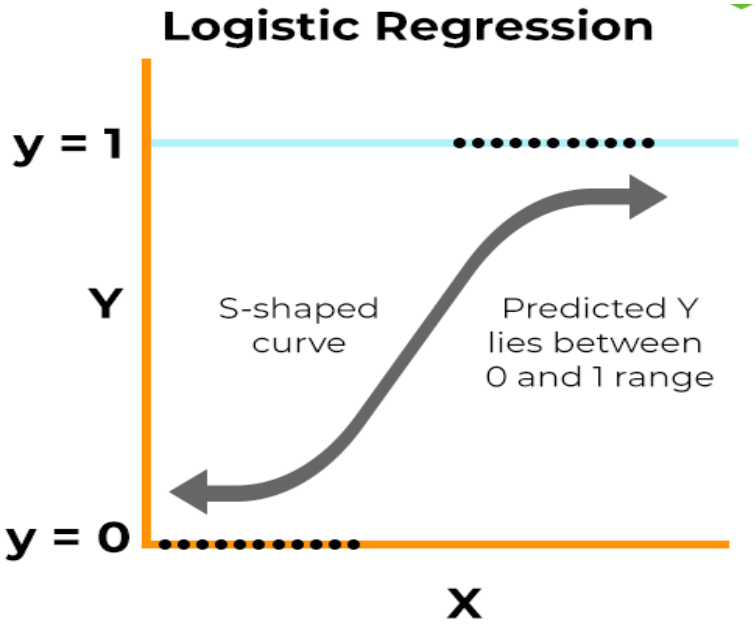
Batch : DS-C70

|

Date : 20-01-2025

What is Logistic Regression?

Logistic regression is a supervised machine learning algorithm used for classification tasks.

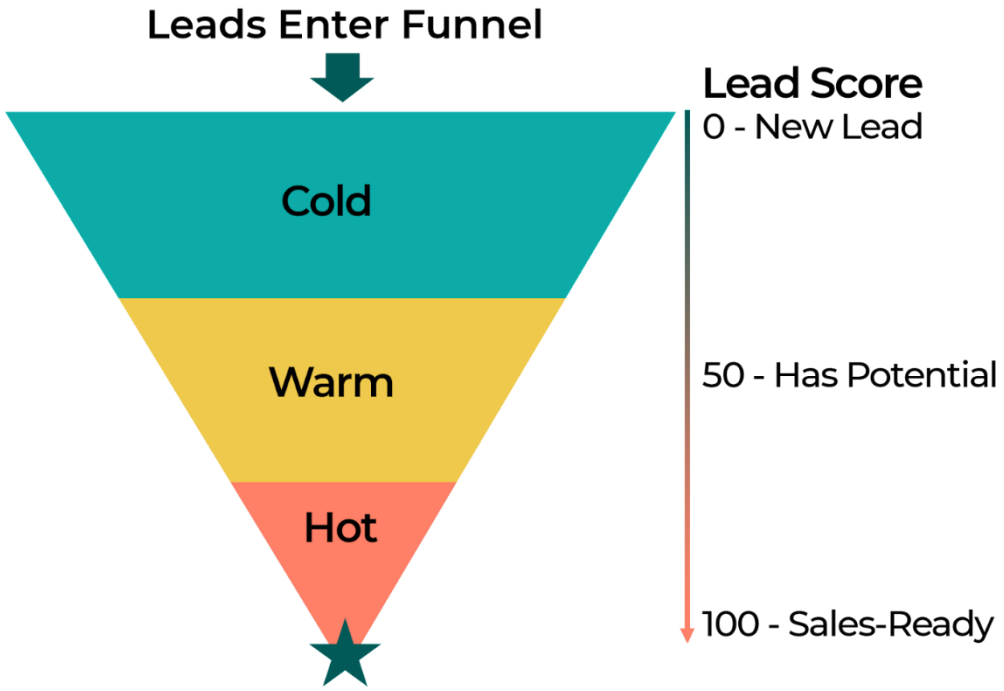


Problem Statement :-

- X Education company sells online courses to industry professionals through various marketing channels. But its lead conversion rate is very poor.
- The company wishes to identify the 'Hot Leads' for Marketing Team to focus only on them.

Hot Lead Funnel Structure

Please find below image to explain the concept of Hot Leads Demonstrated as a Funnel.



Basic steps of Logistic Regression in a Pipeline format :

Please find below image to explain the basic steps of a Logistic Regression in step by step self-explanatory pipeline format.

Steps followed During Building the Model

The following are the steps used:

1. Data Understanding: - Load the data, create the dataset and understand the meaning of the dataset. Check for data imbalance, null values, NULL values with SELECT word mentioned as no values has been selected in that.

2. Data Cleaning: -

    i. Transformed SELECT Values to NULL
    ii. Checked null values > 40% and removed the columns
    iii. Checked for Unique values to drop the columns
    iv. Imputed rest of the NULL values with UNKNOWN and OTHER

    Finally, the dataset is cleaned and now no NULL Values exist in the dataset

3. Outlier checking: - We saw few outliers in the numerical columns. We have imputed them with Mean as they were very less in number i.e only 1%.

4. EDA: A quick EDA was done to check the condition of our data. Checked both Categorical data, Numerical data as well as distribution for our target variable 'Converted'. The numeric values seems good and no outliers were found as we have imputed the outliers in the previous step.

5. Data Transformation: Changed the categorical binary variables into '0' and '1' for ease of Model Building.

6. Dummy Variables Creation:

    i. We created dummy variables for the categorical variables.
    ii. Removed all the repeated and redundant variables.

    At the end we finally got 155 columns to go with the initial Model.

7. Test-Train Split: The first step for Building a Logistic Regression is performing a train-test split, we have chosen 70:30 ratio. 70% data as Train data and 30% data as Test data.

8. Feature Rescaling:

    i. We used the Standard Scaling from sklearn package and preprocessing Library to scale the all variables.

ii. Then, we plot the heatmap to check the correlations among the features.
iii. NaN producing features has been eliminated

9. Model Building: Using RFE to select top 15 relevant variables. Finally, we got 12 most significant variables. The VIF's and P-values for the last and 5th Model are good. (Considering threshold of VIF < 5 and p-value < 0.05).

10. Model Evaluation:

    i. Built Confusion Matrix.
    ii. Drawn ROC Curve to find the optimum cut off
    iii. Using this we find the accuracy, sensitivity and specificity 80% each.
    iv. Performed Precision and recall tradeoff to get another view of the accuracy

11. Prediction: Prediction was done on the test data frame and with an optimum cut off as 0.29 with accuracy, sensitivity and specificity of almost 80%.

Conclusion:

- The lead score calculated in the test set of data shows the conversion rate of 92% on the final predicted model which clearly meets the expectation of CEO has given a ballpark of the target lead conversion rate to be around 80%.

- Good value of sensitivity of our model will help to select the most promising leads.

- Features which contribute more towards the probability of a lead getting converted are

    i. Tags_CLOSED BY HORIZZON
    ii. Tags_LOST TO EINS
    iii. Tags_WILL REVERT AFTER READING THE EMAIL

Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.