

Summary Report

By: Pratyush Srivastava

Introduction:

X Education approached us with a business problem: their lead conversion rate is about 30%, and they would like to improve it to around 80%. Our task was to analyse historical lead data and build a predictive model to identify high-potential 'hot leads' that the sales team can prioritize.

Approach:

I began by thoroughly examining the provided dataset, identifying missing values, and cleaning up irrelevant categories like 'Select'. After cleaning, I prepared the data by encoding categorical variables and scaling numerical variables. The logistic regression model was chosen for this binary classification problem, as it is interpretable and effective for lead scoring.

The data was split into training and test sets, and the logistic regression model was trained on the training set. Afterward, predictions were made on the test set, and the model's performance was evaluated using metrics such as accuracy, precision, recall, and the ROC-AUC score.

Results:

The model achieved:

- Accuracy: 82.04%
- Precision: 80.30%
- Recall: 71.44%
- ROC-AUC: 90.16%

These metrics indicate that the model is well-calibrated for distinguishing between hot and cold leads. By focusing on the most promising leads (those with high lead scores), the sales team can increase conversion rates and reduce wasted efforts.

Business Recommendations:

With the current model, X Education can now assign lead scores to new leads, and the sales team can prioritize follow-ups with hot leads. This data-driven approach can increase the company's overall lead conversion rate. The model can also be further fine-tuned and expanded to handle changing business requirements.

Learnings:

Throughout the project, we learned the importance of properly cleaning and pre-processing data, especially handling missing values and irrelevant categories. Model interpretability was key, as logistic regression provides insights into how each feature contributes to the conversion probability. This project highlighted the business value of predictive analytics in sales operations.