

Summary Report

The analysis for X Education Company focused on understanding what drives industry professionals to enroll in its courses. By examining user behavior—such as website visits, time spent on the platform, traffic sources, and conversion rates—the goal was to identify key factors influencing enrollment decisions.

The first step was data cleaning, where missing values were addressed by either replacing them with "Not Provided" or removing them when creating dummy variables. The "option select" field, which didn't offer meaningful insights, was discarded. Since a significant portion of the data came from India, the country field was grouped into three categories: India, Outside India, and Not Provided. A quick exploratory data analysis (EDA) revealed that some categorical variables contained irrelevant elements, while numerical data was well-structured with no significant outliers.

To prepare the data for modeling, categorical variables were transformed into dummy variables, and numerical values were normalized using the MinMaxScaler. The dataset was then split into 70% training data and 30% test data. For feature selection, Recursive Feature Elimination (RFE) helped identify the 15 most relevant variables. Additional refinements were made by eliminating variables based on Variance Inflation Factor ($VIF < 5$) and p-values (< 0.05) to improve model accuracy.

Model evaluation involved using a confusion matrix, while an ROC curve helped determine an optimal cutoff value of 0.35, leading to approximately 80% accuracy, sensitivity, and specificity. Testing the model on the test dataset confirmed similar results. A Precision-Recall Curve was also analyzed, identifying an optimal cutoff of 0.41, with 73% precision and 75% recall.

The analysis highlighted the most influential factors in lead conversion: total time spent on the website, lead origin (Lead Add Form), current occupation (Working Professional), lead source (Welingak Website), and total visits. These insights suggest that X Education can improve its conversion rates by enhancing website engagement, refining lead capture forms, and targeting working professionals through more strategic marketing efforts.