Lead Scoring Case Study

Nitesh Pradhan

Project Overview

- Objective: Identify leads most likely to convert
- Business context: EdTech company with multiple lead sources
- Approach: Build a logistic regression-based predictive model

Step 1: Data Cleaning

- Dropped columns with >3000 missing values
- Removed irrelevant features (e.g., 'City')
- Filled remaining nulls using appropriate techniques

Step 2: Exploratory Data Analysis

- Analyzed distribution of categorical and numerical features
- Plotted lead source, last activity, and interest levels
- Identified key features influencing lead conversion

Step 3: Feature Engineering

- Created dummy variables for categorical features
- Standardized numeric features where needed
- Prepared data for modeling

Step 4: Model Building

- Used logistic regression for classification
- Feature selection with Recursive Feature Elimination (RFE)
- Split data into train-test sets

Step 5: Model Evaluation

- Evaluated using accuracy, precision, recall, F1 score
- Plotted ROC curve, computed AUC
- Selected best threshold for classification

Business Impact

- Model helps prioritize high-quality leads
- Supports better resource allocation for sales teams
- Improves overall lead conversion rate