



LEAD SCORING CASE STUDY ASSIGNMENT

Submitted by : Ankit Prakash



PROBLEM STATEMENT

- **Company Background:**
 - X Education sells online courses to industry professionals through its website and other marketing channels (e.g., Google, referrals).
- **Lead Generation & Conversion Challenge:**
 - Many visitors show interest by filling out a form, becoming "leads."
 - Sales team follows up with these leads via calls, emails, etc., but the **current lead conversion rate is only 30%**.
- **Objective:**
 - **Increase Lead Conversion Efficiency** by identifying "Hot Leads" (leads with high potential to convert).
 - Focus sales efforts on high-potential leads to boost conversion rate.
- **Goal:**
 - Develop a **Logistic Regression Model** to assign a **Lead Score (0-100)**, predicting the likelihood of each lead converting.
 - Target a **lead conversion rate of 80%** for high-scoring leads.
- **Data Overview:**
 - Dataset of around 9,000 leads with features like Lead Source, Total Visits, Time Spent on Website, etc.
 - Target variable: 'Converted' (1 = converted, 0 = not converted).
 - Categorical values like "Select" need handling as they indicate missing data.



ANALYSIS AND MODEL BUILDING STEPS

- **Data Import & Inspection**
 - Load data and inspect the structure, data types, and missing values.
- **Data Preparation**
 - Clean and preprocess data, handle missing values and outliers.
- **Exploratory Data Analysis (EDA)**
 - Analyze data distributions, trends, and key insights related to lead conversion.
- **Dummy Variable Creation**
 - Convert categorical variables into numerical dummy variables for modeling.
- **Train-Test Split**
 - Split data into training and test sets for model evaluation.
- **Feature Scaling**
 - Standardize numerical features to improve model performance.
- **Correlation Analysis**
 - Check feature correlations to identify and manage multicollinearity.
- **Model Building**
 - Use **Recursive Feature Elimination (RFE)** for feature selection.
 - Evaluate **R-squared**, **VIF**, and **p-values** to select optimal features.
- **Model Evaluation**
 - Assess model performance on test data using metrics such as accuracy, precision



LEAD CONVERSION ANALYSIS AND PREDICTIVE MODELING

- **Objective:** Develop a data-driven model to predict lead conversion probability.**Data Preparation:****Missing Values:** Handled with imputation methods where necessary.
- **Data Transformation:** Converted categorical data using **One-Hot Encoding**.
- **Data Scaling:** Applied **StandardScaler** for numeric columns to ensure uniformity.
- **Analysis Goal:** Identify the most impactful features on lead conversion, creating a streamlined, actionable model for business insights.



FEATURE SELECTION WITH RECURSIVE FEATURE ELIMINATION (RFE)

○ Feature Selection:

- **RFE Method:** Selected top 15 features using Recursive Feature Elimination with Logistic Regression.
- **Key Features:** Total Visits, Lead Source, Time Spent on Website, etc.

○ Model Building - Logistic Regression:

- **Metrics Used:**
 - **P-Values:** Determined feature significance.
 - **VIF (Variance Inflation Factor):** Checked for multicollinearity.
- **Findings:**
 - Refined features by removing insignificant ones (high p-values) and those with multicollinearity (high VIFs).
- **Outcome:** Achieved a refined model focusing on the most relevant predictors.



MODEL INSIGHTS AND STRATEGIC RECOMMENDATIONS

- **Significant Predictors:** **Total Visits & Time Spent on Website:** Strong positive correlation with conversion likelihood.
- **Lead Source** (e.g., **Reference** and **Direct Traffic**): Differing impact on conversion rates.
- **Model Interpretation:** Focused marketing on sources with high conversion potential.
- Enhanced engagement strategies to boost website interaction metrics.
- **Business Takeaways:** Use insights to prioritize engagement initiatives and optimize marketing spend on channels with higher conversion rates.



THANK YOU...!!!

