**Name: -** Vatsal Rameshbhai Gohel

**Mail id: -** [**gohel.vatsal11@gmail.com**](mailto:gohel.vatsal11@gmail.com)

**Contact Detail: -** +1-774-253-3960

**LinkedIn Id: -** <https://www.linkedin.com/in/gohelvatsal/>

**GitHub Id: -** <https://github.com/vatsalgohel>

**Project Portfolio**

**Lead Scoring for X Education**

**Project Title:**

Enhancing Lead Conversion Through Predictive Modeling

**Problem Statement:**

X Education, an online course provider, faces challenges with low lead conversion rates despite acquiring a substantial number of leads daily. The company aims to improve efficiency by identifying and prioritizing high-potential leads, referred to as 'Hot Leads', using predictive modeling techniques.

**Project Description:**

**Objective:**

The objective of this project is to build a predictive model that assigns a lead score to potential customers based on their likelihood to convert into paying customers. This model aims to increase lead conversion rates from the current 30% to approximately 80%.

**Scope:**

The scope includes sourcing and cleaning the leads dataset, performing exploratory data analysis (EDA), building a logistic regression model for lead scoring, evaluating the model's performance metrics, and recommending the deployment of the model for future lead prioritization.

**Deliverables:**

* Cleaned and prepared dataset for analysis.
* Logistic regression model for lead scoring.
* Evaluation metrics including accuracy, sensitivity, specificity, precision, and recall.
* Recommendations for using the model to prioritize leads effectively.

**Methodology:**

**Tools and Technologies Used:**

* **Programming Language:** Python
* **Libraries:** pandas, NumPy, scikit-learn, stats models
* **Data Visualization:** matplotlib, seaborn
* **Database:** Excel file for data storage and retrieval
* **Documentation:** Microsoft Word for report writing, PowerPoint for presentation

**Process:**

* **Data Collection and Cleaning:**
  + Importing the dataset and handling missing values.
  + Removing irrelevant columns and handling 'Select' options.
* **Exploratory Data Analysis (EDA):**
  + Analysing distributions and relationships within the dataset.
  + Identifying key features correlated with lead conversion.
* **Feature Engineering:**
  + Creating dummy variables for categorical features.
  + Scaling numerical features using StandardScaler.
* **Model Building:**
  + Splitting the dataset into training and testing sets.
  + Building a logistic regression model to predict lead conversion.
  + Using Recursive Feature Elimination (RFE) for feature selection.
* **Model Evaluation:**
  + Assessing model performance using metrics such as accuracy, sensitivity, specificity, precision, and recall.
  + Plotting ROC curves and determining optimal cutoff points.
* **Deployment and Recommendations:**
  + Deploying the final model to predict lead scores in real-time.
  + Providing recommendations to X Education on utilizing the model for prioritizing leads effectively.

**Role and Responsibilities:**

* **User:** Lead Data Analyst
  + Managed overall project execution, including data sourcing, cleaning, analysis, and modeling.
  + Responsible for ensuring the accuracy and reliability of model predictions.
  + Collaborated with team members to finalize reports and presentations.

**Results and Impact:**

**Outcome:**

The logistic regression model achieved an accuracy of 79.1%, sensitivity of 74.3%, and specificity of 81.8% in predicting lead conversions on the test dataset.

**Impact:**

* Increased efficiency in lead prioritization, potentially raising lead conversion rates to 80%.
* Provided actionable insights into factors influencing lead conversions, such as lead source and engagement metrics.

**Challenges and Solutions:**

**Challenges Faced:**

* Handling 'Select' options in categorical variables.
* Dealing with imbalanced datasets and optimizing model performance.

**Solutions Implemented:**

* Replaced 'Select' options with appropriate null values.
* Used techniques like feature scaling and RFE for model optimization.

**Conclusion and Learnings:**

**Summary:**

The project successfully developed a predictive model for lead scoring, addressing X Education's challenge of improving lead conversion rates through data-driven insights.

**Learning Experience:**

* Enhanced proficiency in data preprocessing, exploratory analysis, and predictive modeling.
* Gained insights into the significance of feature selection and model evaluation in achieving business objectives.

**Link for the file:**

<https://drive.google.com/drive/folders/1OwlPmjpW0BKZZia14iKU5zGozuUFN2s3?usp=sharing>