***Propensify: Propensity Model for Marketing Campaign Optimization***

***Overview***

**Propensify** is a project designed to build a propensity model to identify how likely certain target groups of customers are to respond to a marketing campaign. The goal is to optimize marketing efforts by predicting which potential customers are most likely to engage.

***Problem Statement***

Many businesses invest heavily in data-driven marketing campaigns with the hope of predicting customer behaviour. However, they often struggle to achieve meaningful results. This project aims to leverage historical data to build a model that forecasts customer responses, helping businesses make more informed marketing decisions.

***Data***

The dataset provided by the insurance company includes:

* train.csv: Historical data for training the model.
* test.csv: List of potential customers to assess.

**Note:** Ignore any columns not listed in the provided dataset description.

***Project Steps***

1. **Exploratory Data Analysis (EDA)**
   * Analyze the data to identify patterns, relationships, and trends.
   * Use descriptive statistics and visualizations.
2. **Data Cleaning**
   * Standardize data and handle missing values and outliers.
3. **Handling Imbalanced Data**
   * Address the imbalance in the dataset using appropriate techniques.
4. **Feature Engineering**
   * Create or transform features to enhance model performance.
5. **Model Selection**
   * Choose suitable models for classification tasks.
6. **Model Training**
   * Split the data into training and testing sets.
   * Train models to find optimal parameters.
7. **Model Validation**
   * Evaluate model performance on unseen data to check generalization and identify overfitting.
8. **Model Deployment**
   * Prepare the trained model for use in a production environment.

***Timeline***

Complete the project within 2 weeks, ensuring all deliverables are met.

***Deliverables***

Submit the following in a zip file:

* **Report (PDF)**
  + Description of design choices
  + Performance evaluation of the model
  + Discussion of future work
* **Source Code**
  + Code used to create the pipeline

***Tasks/Activities List***

* Collect and load data from train.csv and test.csv.
* Perform EDA: Data quality checks, missing values, and outlier treatment.
* Ensure correct data types, especially for date fields.
* Balance the dataset.
* Feature engineering and selection.
* Split data into train and test sets; apply sampling to find optimal splits.
* Choose metrics for model evaluation.
* Test multiple classification models and select the best one.
* Model selection, training, prediction, and assessment.
* Hyperparameter tuning and model improvement.
* Update test.csv with a new column indicating whether to market to each candidate (1 for yes, 0 for no).
* Create a model deployment plan.

***Success Metrics***

* Achieve model accuracy on the test dataset > 85%.
* Implement hyperparameter tuning methods.
* Perform thorough model validation.

***Bonus Points***

* Package the solution with a README.md explaining installation and execution of the pipeline.
* Demonstrate documentation skills, highlighting how the solution benefits the company.

***Data***

Access the dataset through the provided link.

