Donor Retention Prediction

Problem Statement:

In this problem, you are working for a nonprofit organization that depends heavily on recurring donations for its sustainability. Retaining donors over the long term is essential, but the nonprofit is struggling to do so. Your goal is to predict which donors are at risk of discontinuing their donations (churn) based on historical donation data and engagement records. This organization has limited technical resources, so the model must prioritize efficiency, clarity, and the ability to drive actionable outcomes.

Engagement metrics such as event participation, newsletter openings, and email clicks are good indicators of donor involvement, while donation frequency and demographics like age and income provide further context. Nonprofits must comply with privacy regulations like HIPPA, so handling personal data with care is crucial in this analysis.

Finally, the goal isn't just to predict churn but to align the model with real-world retention strategies. Consider how your model's predictions could be used by a nonprofit with limited resources to develop targeted, personalized retention strategies.

Data Provided:

- **Historical donation data** (amounts, frequency, donation dates)
- **Demographic information** (age, location, income range, etc.)
- **Engagement records** (participation in events, newsletter opens, email clicks)
- **Communication preferences** (preferred channels, frequency of communications)

Machine Learning Task:

- Develop a model to predict whether a donor will continue donating in the next cycle or churn.

Deliverables:

In this case study, the focus is on how you **map out and communicate** your problem-solving approach. Your ability to explain your thought process, use diagrams, and provide clear written responses is key. Here are the main steps:

1. Data Understanding

- **Task**: Explore the dataset, identify key features, and note any data quality issues (e.g., missing values, outliers).

- **Objective**: Address how you plan to clean and prepare the data for analysis. Highlight significant insights about the dataset. Max 200 words.

2. Modeling Strategy

- **Task**: Select and justify a machine learning model that aligns with the business goal of predicting donor churn.
- **Objective**: Explain why you chose this model and how it fits the organization's needs, balancing performance and simplicity. Max 200 words.

3. Process Mapping

- **Task**: Outline the entire process from data preparation to model evaluation and deployment.
- **Objective**: Describe the key steps in building the solution, emphasizing efficiency and effectiveness. Max 100 words. A diagram or flowchart is required to illustrate your process.

Format of Deliverables:

- For deliverables 1 and 2, please use Jupyter notebooks (markdown file) or similar tool for code and written insights.
- For deliverable 3, please use word, excel, or other process mapping tools for diagram or flowchart.

Evaluation Criteria:

- **Technical Efficiency**: How well you handle data preprocessing, feature engineering, and model training, especially with messy data.
- **Resourcefulness**: Your ability to work with limited resources and open-source tools.
- **Business Alignment**: How well your solution ties into donor retention strategies and respects privacy regulations.
- **Model Performance**: Your rationale for model selection and evaluation metrics (e.g., accuracy, precision, recall), and how you manage imbalanced data.