AI Task: Customer Churn Prediction

Task Description:

Develop a machine learning model to predict customer churn using the provided dataset. The task involves preprocessing the data, selecting appropriate features, training a model, and evaluating its performance.

Dataset

The dataset provided (`churn-bigml-80.csv`) contains various features related to customer behavior and demographics. The target variable is whether the customer has churned (yes/no).

Requirements:

1. **Data Preprocessing**:
   - Load the data and check for any missing values.
   - Perform necessary data cleaning and preprocessing.
   - Convert categorical variables to numerical format.
   - Scale or normalize features as needed.
2. **Exploratory Data Analysis (EDA)**:
   - Conduct EDA to understand the data distribution and relationships between features.
   - Visualize key insights using appropriate plots and charts.
3. **Model Development**:
   - Split the data into training and testing sets.
   - Train a machine learning model to predict customer churn. Consider models like Logistic Regression, Random Forest, or Gradient Boosting.
   - Perform hyperparameter tuning to optimize the model.
4. **Model Evaluation**:
   - Evaluate the model using metrics such as accuracy, precision, recall, F1 score, and ROC-AUC.
   - Generate and interpret a confusion matrix.
5. **Documentation**:
   - Document the entire process, including data preprocessing, model training, and evaluation.
   - Provide insights and observations based on your findings.
6. **Submission**:
   - Submit a Jupyter Notebook (.ipynb) that includes all the steps from data preprocessing to model evaluation.
   - Submit a Python script (.py) that contains the final version of the code for data preprocessing, model training, and evaluation.

- Include a `README.md` file with instructions on how to set up and run the project.

Submission Details:

- **Repository**: Submit the project via a public GitHub repository.
- **Naming Structure**: `Firstname-Lastname-ChurnPrediction-Task`
- **Deadline**: Please submit your completed task by July 18, 2024.
- **Submission Email**: Send the link to your public GitHub repository to samprad@themakercrew.com with the subject line "{Firstname} {Lastname} Churn Prediction Task".

Evaluation Criteria:

1. **Functionality**: Does the model correctly predict customer churn?
2. **Code Quality**: Is the code clear, modular, and well-documented?
3. **Model Performance**: How accurate and reliable are the churn predictions?
4. **Data Preprocessing**: How effectively is the data cleaned and prepared?
5. **Model Evaluation**: Are the evaluation metrics and insights thorough and well-explained?
6. **Documentation**: Is the setup and usage documentation clear and comprehensive?

Note:

If you have any questions or need further clarification, please do not hesitate to reach out at samprad@themakercrew.com. We are here to help ensure you understand the requirements and can complete the task successfully