Machine Learning Project

Project Title : Customer Churn Prediction

Assignment Description:

At Sunbase, we prioritize understanding our customers and ensuring their satisfaction. To achieve this, we want to develop a machine learning model that predicts customer churn. Your task as a Machine Learning Intern is to work on this project, following the guidelines and responsibilities outlined in the job description.

Objective:

Develop a machine learning model to predict customer churn based on historical customer data. You will follow a typical machine learning project pipeline, from data preprocessing to model deployment.

Data:

You are provided with a dataset containing historical customer information, including customer attributes, interactions, and whether they churned or not. The dataset is in CSV format.

Tasks:

1. Data Preprocessing:

- Load the provided dataset and perform initial data exploration.
- Handle missing data and outliers.
- Prepare the data for machine learning by encoding categorical variables and splitting it into training and testing sets.

2. Feature Engineering:

- Generate relevant features from the dataset that can help improve the model's prediction accuracy.
- Apply feature scaling or normalization if necessary.

3. Model Building:

- Choose appropriate machine learning algorithms (e.g., logistic regression, random forest, or neural networks).
- Train and validate the selected model on the training dataset.
- Evaluate the model's performance using appropriate metrics (e.g., accuracy, precision, recall, F1-score).

4. Model Optimization:

- Fine-tune the model parameters to improve its predictive performance.
- Explore techniques like cross-validation and hyperparameter tuning.

5. Model Deployment:

- Once satisfied with the model's performance, deploy it into a production-like environment (you can simulate this in a development environment).
- Ensure the model can take new customer data as input and provide churn predictions.

Deliverables:

1. Jupyter Notebook or Python script containing your code.
2. A brief report summarizing your approach, including any data preprocessing, feature engineering, and model selection decisions.
3. Model performance metrics and visualizations.

Additional Information:

- You are encouraged to use Python and relevant machine learning libraries like scikit-learn, TensorFlow, or PyTorch.
- We expect you to demonstrate your skills in data preprocessing, feature engineering, model building, and deployment.

Evaluation Criteria:

Your assessment will be evaluated based on the following criteria:

- Data preprocessing and cleaning.
- Feature engineering creativity and effectiveness.
- Model selection and optimization.
- Model deployment and integration.
- Clarity and organization of your code.
- Documentation and reporting of your work.

Note:

This assignment is designed to evaluate your skills and approach to a real-world machine learning problem, aligning with the responsibilities of a Machine Learning Engineer at Sunbase. Good luck with your project, and we look forward to reviewing your work!