Atidot

Exercise: Predicting Customer Churn Using Time-Series Data

Objective: Build a **robust** classification model that predicts which customers will churn between **Jan. 1, 2024** and **Feb. 28, 2024**, using **time-series data** with a double index (customer_id and date). Demonstrate modeling and feature engineering skills, as well as production-aware design practices.

Deliverables: Produce a GitHub repository named customer-churn-classification that contains:

1. A README.md file to shortly explain the purpose, setup, and usage of the repository.
2. The fully functional code in clearly commented **.py** scripts (no notebook).
3. The original **.csv** file, with an additional prediction column appended.
4. A **.json** file containing precision, recall, and at least one additional relevant metric.
5. A trained model artifact (e.g., a **.pkl** file).

Estimated time to completion: up to 120 minutes of focused work time. The emphasis is on **quality over completeness**. If you can't implement everything perfectly within that time, explain how you'd **prioritize** and **extend** your solution in a real-world scenario.

Data Overview: The dataset (churn_data.csv) contains a double index (customer_id, date), several features, and a **target variable** (churn, which is binary: 1 for churned, 0 otherwise). The data has been synthetically generated but should be treated as real-world data.

Instructions:

1. **Data Preprocessing:**
    - Handle missing values.
    - Appropriately encode (1-hot, embeddings, etc.) categorical features.
2. **Data Engineering:**
    - Generate at least 3 relevant date-dependent features.
    - Enrich the data with one relevant external source (e.g., economic indicators).
3. **Modeling:**
    - Select an appropriate algorithm for churn classification, and implement it. Make sure to systematically tune and validate your model.
    - Be aware of data leakage, and provide a clear rationale for your choices.
4. **Model explanation:**
    - Use an appropriate method for **brief** model explanation (e.g., SHAP, PDP, LIME).
    - Shortly describe why the method you chose is technically relevant**.**