Project

Customer Churn Prediction System: From Data to Deployment

CONTEXT & OBJECTIVE

Telecommunication and subscription-based companies lose significant revenue when customers unsubscribe from their services. Predicting churn allows proactive actions like special offers or engagement campaigns.

Your task is to build an **end-to-end ML pipeline** that predicts whether a customer is likely to churn based on their usage patterns and service history.

DATASET

Use the **Telco Customer Churn Dataset** (available on Kaggle):
☞ https://www.kaggle.com/datasets/blastchar/telco-customer-churn

This dataset includes customer demographics, account information, and usage data with the target column `churn` (Yes/No).

PROJECT GOALS

1. **Data Science Layer:**
   - Perform EDA and visualize key correlations with churn.
   - Handle missing values and categorical encoding.
   - Train and compare multiple binary classifiers (Logistic Regression, Random Forest, SVM, etc.).
   - Optimize performance using cross-validation and hyper parameter tuning.
2. **MLOps Layer:**
   - Modularize code: **`data_preprocessing.py`, `train_model.py`, and `predict_api.py`.**
   - Build a REST API (flask, spring boot, etc…)
   - Containerize all components with Docker and orchestrate with **`docker-compose`.**
   - Visualize API model metrics and performance

EXPECTED DELIVERABLES

- Code repository with modular Python scripts and Docker setup.
- `docker-compose` to run preprocessing, model, API,….
- Report including data analysis, ML workflow, and architecture diagram.
- Dashboard screenshots showing metrics
- Final presentation and defense demonstrating deployment

SUGGESTED TECHNOLOGIES

- **Python Libraries:** pandas, scikit-learn, joblib, FastAPI, prometheus_client
- **MLOps Stack:** Docker, docker-compose, GitHub Actions,
- **Version Control:** Git / GitHub
- **Project management tool** : Trello, ClickUp,…