END-TO-END MLOPS PROJECT: CUSTOMER CHURN PREDICTION

Problem Statement:

Developing an end-to-end MLOps pipeline for Customer Churn Prediction, ensuring continuous monitoring and periodic retraining of the model to maintain accuracy.

Data Source and Data Preprocessing:

- Kaggle Dataset used for customer churn prediction.
- Data preprocessing performed before training.
- Data Preprocessing (Cleaning, Feature Engineering).

Machine Learning Pipeline:

- Model Training & Comparison (Random Forest, XGBoost).
- MLflow Tracking to log metrics & select the best model.
- FastAPI Deployment for real-time predictions.

MLOps Integration using Airflow:

- Airflow DAGs automate the pipeline, ensuring seamless data processing.
- Task Orchestration:
  - Runs preprocessing, drift detection, and retraining at scheduled intervals.
  - Ensures that the model is automatically retrained if drift is detected.

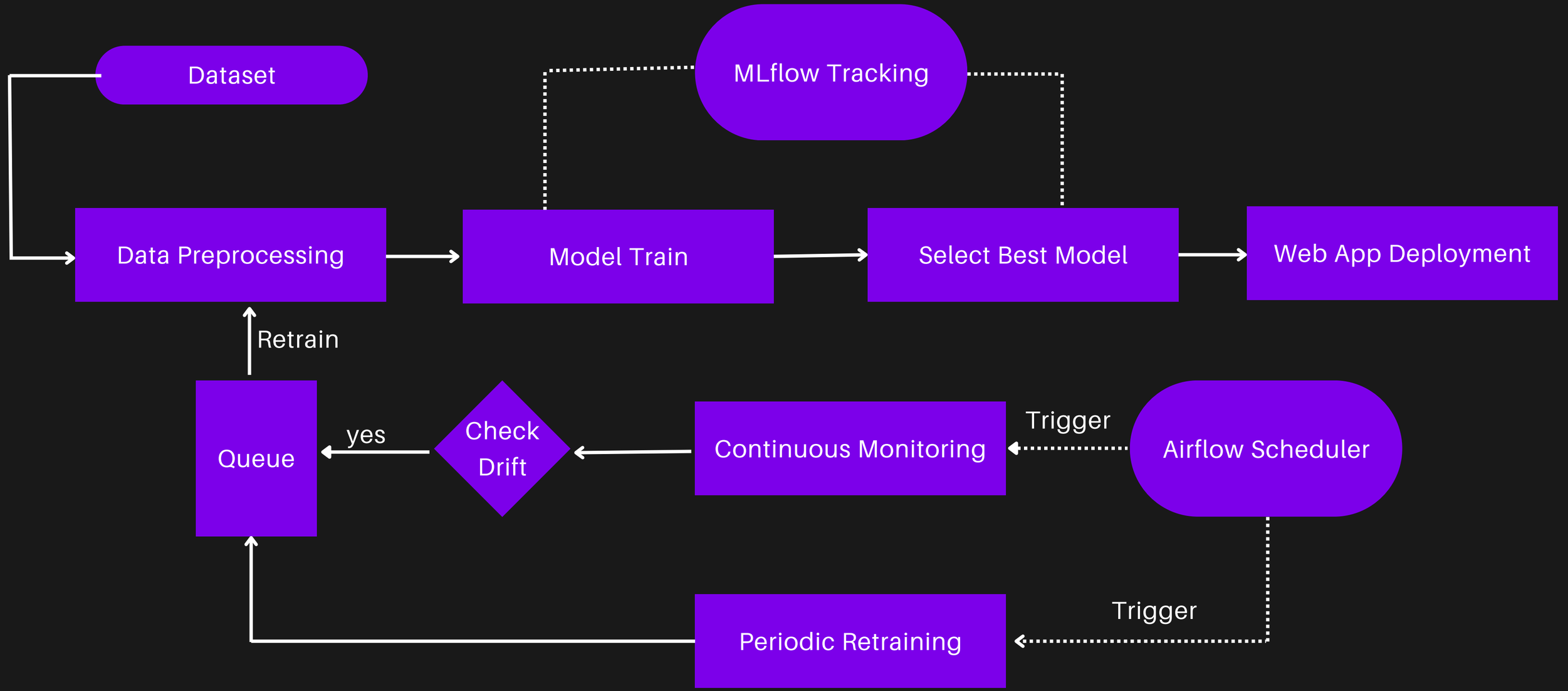
Drift Detection & Retraining:

- Great Expectations used to monitor data drift. If drift is detected, Airflow triggers the retraining pipeline.
- MLflow tracks model performance metrics after each retraining, enabling continuous monitoring and comparison of different model versions.

Deployment & Access:

- Containerized using Docker.
- Deployed on Azure Container Instances for scalable cloud execution.
- Access the deployed model via FastAPI for real-time predictions.

WORK FLOW-


Dataset
MLflow Tracking
Data Preprocessing
Model Train
Select Best Model
Web App Deployment
Retrain
Queue
yes
Check Drift
Continuous Monitoring
Trigger
Airflow Scheduler
Periodic Retraining
Trigger