COMS 6156 - Topics in Software Engineering
Spring 2024 - Revised Project Proposal
Mike Segal | ms6135

Enhancing MLOps with Kubeflow, GitHub Actions, AutoML, and Shap

Introduction

MLOps, the incorporation of DevOps practices into the Machine Learning cycle, is becoming increasingly important as ML is adopted by a growing range of industries. The standardization of MLOps processes and methodologies will enable engineering teams to develop, deploy, and maintain ML services more efficiently and effectively. In order to demonstrate the capabilities of MLOps in creating an automated Machine Learning pipeline, we intend to extend the functionalities of Kubeflow by incorporating GitHub Actions for automated pipeline executions and leveraging AutoML features for enhancing model performance and development velocity. While related works have examined parts of this pipeline (such as GitHub Actions with Kubeflow or Kuberflow with AutoML), we intend to close this gap by fully implementing the pipeline and conducting thorough statistical analysis of each phase.

Project Topic and Motivation

The proposed project focuses on developing an innovative system that integrates GitHub Actions with Kubeflow pipelines, facilitating automated trigger of machine learning workflows with GitHub commits or with scheduled data ingestion. Additionally, the project will explore the impact of employing Katib's AutoML features within Kubeflow pipelines, including hyperparameter tuning, early stopping, and Neural Architecture Search (NAS), on model development and performance.

By automating workflow triggers and incorporating advanced AutoML capabilities, the project aims to significantly reduce the time and effort required to identify optimal model configurations, thereby accelerating the overall MLOps cycle. Data quality checks and model interpretability components will also enhance the quality and transparency of the models.

Objectives

1. To develop a system that integrates GitHub Actions with Kubeflow pipelines, enabling automated execution of machine learning workflows upon GitHub commits.
2. To empirically evaluate the ability of Kubeflow to preprocess datasets and train and evaluate Machine Learning models across various modalities, including Computer Vision and Natural Language Processing.
3. To assess the impact of integrating Katib's AutoML features into Kubeflow pipelines on development velocity and model performance, with a focus on hyperparameter tuning, early stopping, and Neural Architecture Search.
4. To incorporate a data validation tool to maintain high quality data for the ML training phase, as well as a model interpretability functionality as part of the pipeline to ensure greater transparency.

Methodology

1. GitHub Actions will be integrated with Kubeflow Pipelines, so that Kubeflow pipelines runs can be initiated with GitHub commits.
2. Experiments will be conducted to preprocess datasets and train and evaluate models using Kubeflow pipelines across different modalities. Publicly available datasets with known benchmarks will be used to facilitate comparison and reproducibility.
3. Data quality testing using SodaCore will be conducted to ensure all data is suitable for model training.

4. To assess MLOps applicability to Computer Vision, Convolutional Neural Networks for image classification will be developed on the ImageNet dataset.
5. For Natural Language Processing, Recurrent Neural Networks will be employed for text classification on the GLUE or SQuAD datasets.
6. Comparative analysis will be conducted comparing the ML hyperparameter tuning process by hand vs. AutoML, and each modality will be evaluated for its compatibility with the automated MLOps pipeline.
7. Statistical analysis will be conducted of the accuracy, precision, recall, F1 score for classification tasks, and mean squared error (MSE) or mean absolute error (MAE) for regression tasks. These metrics will provide a comprehensive view of model performance.
8. To compare the performance of models tuned by Katib's AutoML features versus manual tuning, and to determine if differences in performance metrics are statistically significant, the paired t-test will be employed.
9. Beyond performance, the efficiency of model tuning will be evaluated in terms of time to achieve optimal performance and computational resources used.
10. Model interpretability will be determined for each modality using Shap, and will be deployed as a monitoring component of the pipeline.

The overall project workflow will be:

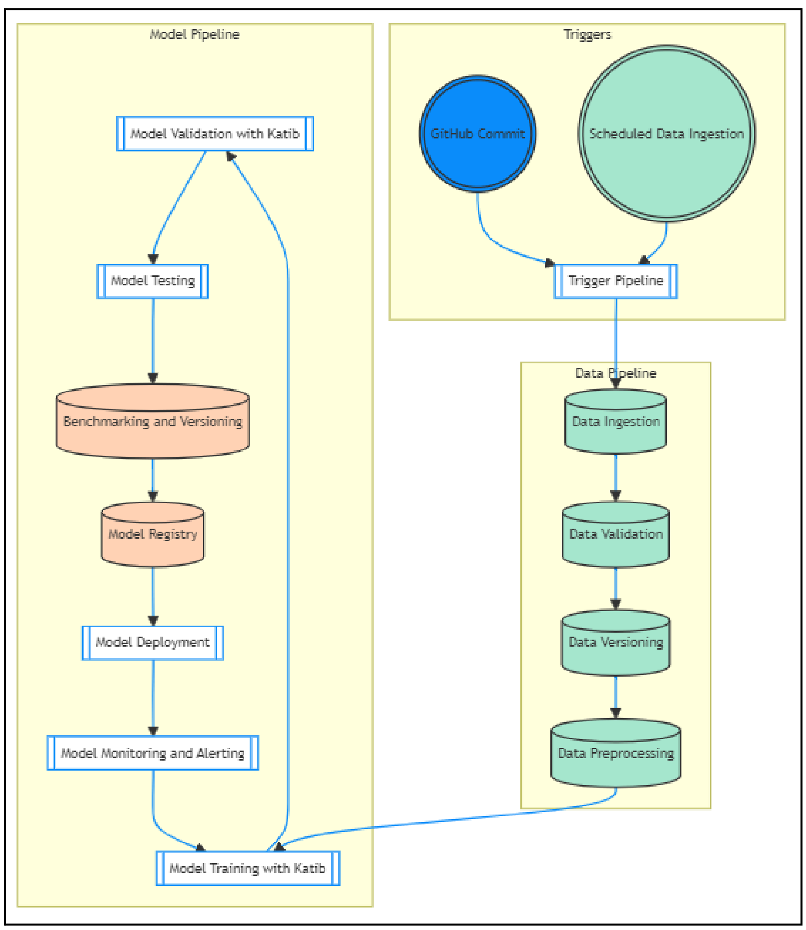

Fig. 1: Proposed Pipeline Workflow

Expected Outcomes

1. A fully functional MLOps pipeline that integrates GitHub Actions with Kubeflow for automated machine learning workflow executions.
2. The incorporation of data quality testing and model interpretability as components of the pipeline.
3. Detailed statistical analysis and performance benchmarks for the MLOps pipeline with and without Katib for multiple datasets and modalities.
4. Clear and rigorous project documentation for future replication and adaptation.

References

https://github.com/sodadata/soda-core
https://github.com/shap/shap