

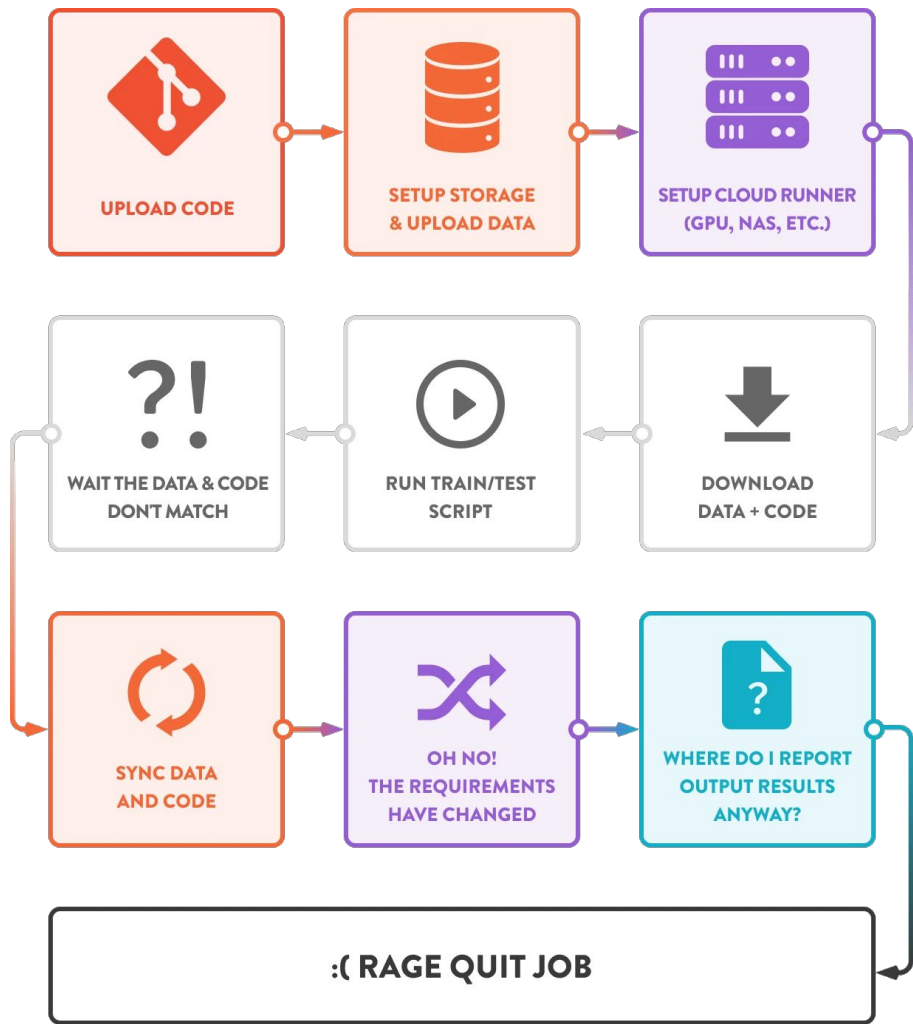
CI/CD for ML

Alex Kim @alex000kim 

Introduction

- Common problems with ML workflows
- Key Assumption about ML workflows
- What's CML (Continuous Machine Learning)?
 - Automated Reporting
 - Infrastructure provisioning
- Live Demo
- Summary

Common problems with ML workflows



Key Assumption:

ML workflows

≈

SWE workflows

Best practices in SWE workflows

- Git repo is source of truth
- Experiment/feature branches
- Automated Reporting
- MRs/PRs and Code Reviews

What's CML (Continuous Machine Learning)?

- Open-source command line tool
- Intended to work within CI tools
 - GitHub Actions
 - GitLab CI/CD
 - Bitbucket Cloud
- Enables GitFlow for ML
- Auto-reporting for PRs/MRs
- Integration with Cloud Services



Create CI workflow

```
name: train-model
on: push
jobs:
  train-model:
    ...
    steps:
      ...
      - name: TrainModel
        ...
        run: |
```

```
  pip install -r requirements.txt
  dvc pull
  dvc repro
  dvc push

  # Create CML report
  echo "## Metrics" >> report.md
  dvc metrics show --md >> report.md
  echo "## Feature Importances" >> report.md
  csv2md reports/feat_imp.csv >> report.md
  echo "## Confusion Matrix" >> report.md
  cml publish reports/figures/cm.png --md >> report.md
  cml send-comment report.md
```

1. Install dependencies and run model training

2. Create custom report

3. Submit report as a comment:
cml comment

View reports in GitLab | GitHub | BitBucket

Changes 1 Pipelines 1

Showing 1 changed file with 2 additions and 1 deletion

Hide whitespace changes Inline Side-by-side

train.py +2 -1 View file @61de689c

```
... @@ -12,7 +12,8 @@ y_test = np.genfromtxt("data/test_labels.csv")
12
13 # Fit a model
14 depth = 10
15 - clf = RandomForestClassifier(max_depth=depth)
16 + n_estimators = 200
17 + clf = RandomForestClassifier(max_depth=depth, n_estimators=n_estimators)
18 clf.fit(X_train, y_train)
19
20 acc = clf.score(X_test, y_test)
...
```

Alexander Kim @alex000kim · 52 minutes ago

Accuracy: 0.872

True label \ Predicted label	0.0	1.0
0.0	0.91	0.087
1.0	0.16	0.84

Need more compute power? CML Runner!

```
name: Run Experiment
on: push
jobs:
  deploy-runner:
    runs-on: ubuntu-latest
    ...
    - name: cml runner
      ...
      run: |
        cml runner \
          --cloud=aws \
          --cloud-region=us-east-1 \
          --cloud-type=g4dn.xlarge \
          --labels=cml-runner

  train-model:
    needs: deploy-runner
    runs-on: [self-hosted, cml-runner]
    container:
      image: iterativeai/cml:0-dvc2-base1
      options: --gpus all
      environment: cloud
    ...
    - name: dvc-repro-cml
      run: |
        # Install dependencies
        pip install -r requirements.txt
        dvc pull
        dvc repro
        dvc push

        ### Create CML report ###
        echo "## Metrics" >> report.md
        dvc metrics show --md >> report.md
        ...
        #####
        cml send-comment --pr --update report.md
```

1. Provision VM in Cloud:
cml runner



2. Install dependencies and
run model training

3. Create custom report

4. Submit report as a
comment

Deploy Runner -> Train Model -> Create Report

Triggered via push last month

alex000kim pushed `cdfda65` `exp`

Status	Total duration	Artifacts
Success	2m 47s	-

1-run-experiments.yaml
on: push

1

2

3

CML PR for exp-tune-hyperparams 192be864 #96

Merged

alex000kim merged 1 commit into `exp-tune-hyperparams` from `exp-tune-hyperparams-cml-pr-192be864` on Sep 22

Conversation 1 | Commits 1 | Checks 0 | Files changed 13 | +102 -102

alex000kim commented on Sep 22

Automated commits for `192be86` created by CML.

CML PR for `192be86` [skip ci] `18a39be`

alex000kim merged commit `bd215e0` into `exp-tune-hyperparams` on Sep 22

alex000kim commented on Sep 22

`18a39be`

Metrics

Path	acc_mean	dice_mean	jacc_mean
metrics.json	0.98028	0.91168	0.83769

A few random test images

exp4_num_155384 Image/True/Pred

CML PR: automatically create a MR (PR)

- create & checkout a new branch
- add files
- commit
- git push
- create a PR
- checkout the original branch

```
$ SHA="$(git log -n1 --format=%h)"
$ BASE="$(git branch)"
$ git checkout -b "${BASE}-cml-pr-${SHA}"
$ git add *.py *.json
$ git commit -m "CML PR for ${SHA} [skip ci]"
$ git push
$ gh pr create -B ${BASE} -t "Suggestions"
$ git checkout -
```

```
- run: |
  echo 'lr = 1.0' >> params.py
  python train.py --output metrics.json
  cml pr *.py *.json
```

Team collaboration:

CML reports become part of your code review process!

Try lightgbm

Edit Code

Open Alexander Kim requested to merge `try-lightgbm` into `master` 2 minutes ago

Overview 1 Commits 2 Pipelines 1 Changes 3

0 0

✓ Pipeline #664257984 passed for 75aad8e on `try-lightgbm` 1 minute ago

Approval is optional

Ready to merge!

Delete source branch Squash commits Edit commit message

1 commit and 1 merge commit will be added to `master` (squashes 2 commits).

Merge

Activity

Sort or filter

Alexander Kim assigned to @alex000kim 2 minutes ago

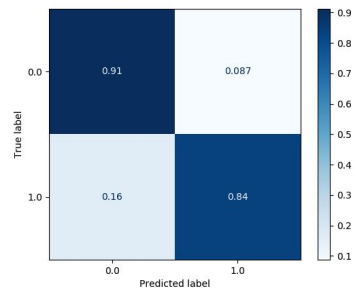
Alexander Kim @alex000kim commented on commit 75aad8e 1 minute ago

Author Owner

Metrics

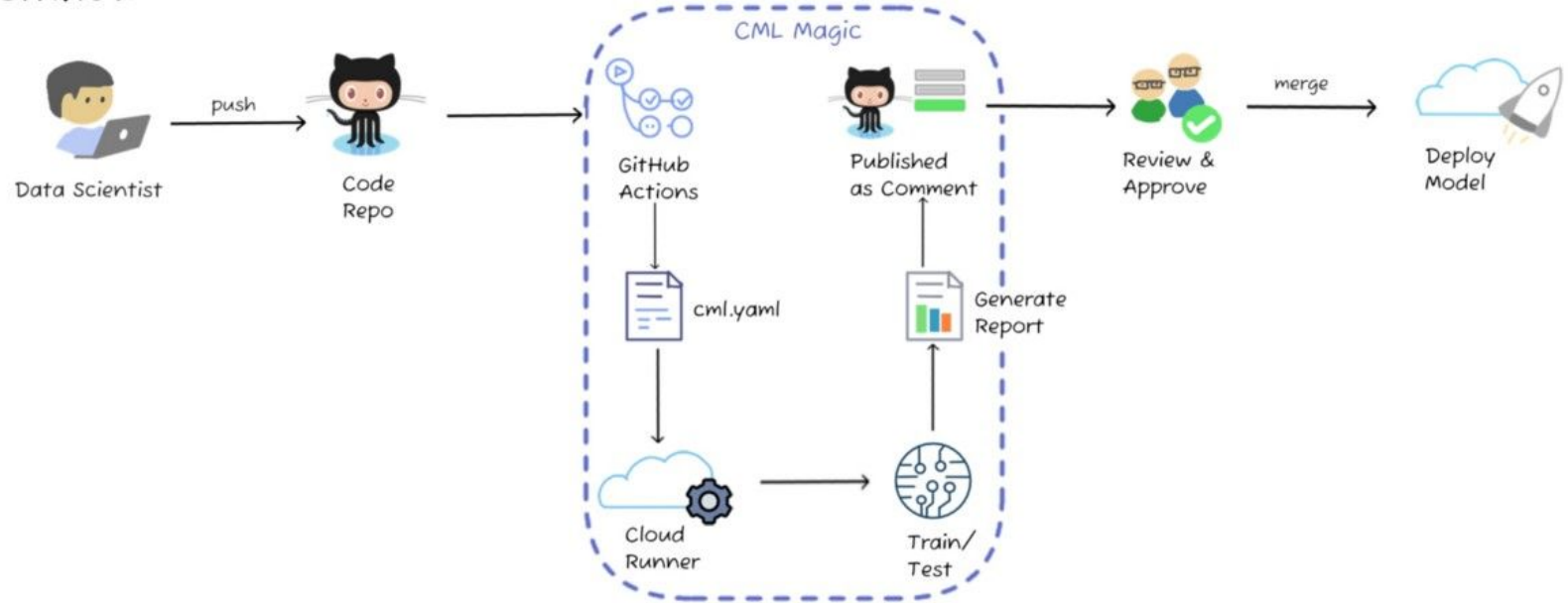
Accuracy: 0.896

Plots



Reply...

CML Workflow

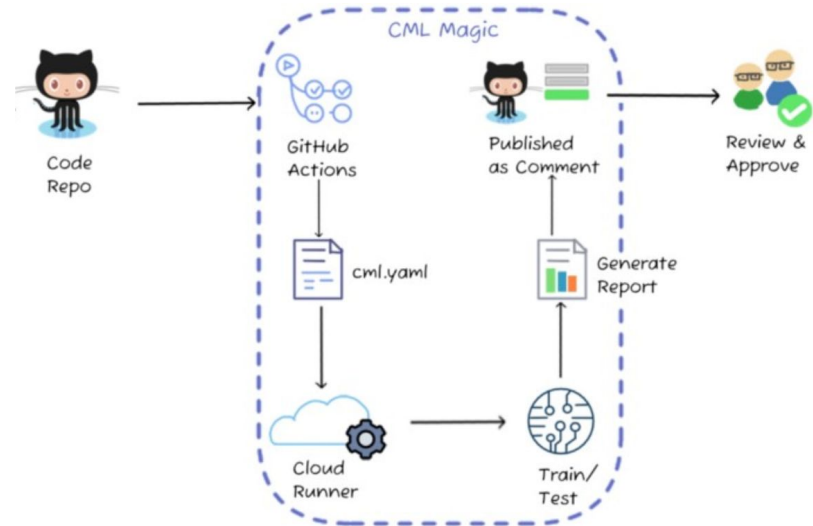


Summary



CML

- ML workflows \approx SWE workflows
- Model training as part of CI/CD
- Automated Reporting
- Easy resource (de)provisioning
- Improved team collaboration!



Practice time!

