

AI4CI

Artificial Intelligence for Continuous integration

Aakanksha Duggal
Hema Veeradhi
Karan Chauhan
Oindrilla Chatterjee
Shrey Anand

Overview

- **What is AI4CI**
- **Tools and Services**
- **Machine Learning Models**
 - **Github Time to Merge**
 - **Optimal Stopping Point**
 - **Build Log Clustering**
- **Engage**

Overview

- **What is AI4CI**
- **Tools and Services**
- **Machine Learning Models**
 - **Github Time to Merge**
 - **Optimal Stopping Point**
 - **Build Log Clustering**
- **Engage**

AIOps

AI + DevOps

Artificial Intelligence for
IT Operations

+

Dev & Ops : **DevOps**

AI Ops

AI + DevOps

CI/CD

Continuous Integration,
Continuous Delivery

Artificial Intelligence for
IT Operations

+

Dev & Ops : **DevOps**

Automating stages of
app development

AI Ops

AI + DevOps

+

=

AI4CI

CI/CD

Continuous Integration,
Continuous Delivery

*AI for Continuous
Integration*

What is AI4CI?



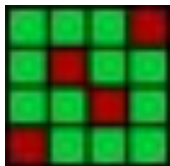
Overview

- What is AI4CI
- **Tools and Services**
- Machine Learning Models
 - Github Time to Merge
 - Optimal Stopping Point
 - Build Log Clustering
- Engage

Tools and Services

CI/CD Data Sources

TestGrid



Prow



Github



AI/ops Tools

Metric
Dashboards

Model Services

ML Services on Operate First

- Tools **built** on Operate First and **available** on Operate First
- AI Services deployed on Operate First Cloud:
 - **Github time-to-merge prediction**
 - **Optimal Stopping point prediction**
 - **Build log clustering**

Overview

- What is AI4CI
- Tools and Services
- Machine Learning Models
 - **Github Time to Merge**
 - Optimal Stopping Point
 - Build Log Clustering
- Engage

GitHub PR TTM: Overview

- How do you calculate, track, and forecast project dev KPIs?
 - velocity, blockers, idea to production time, etc.
- How do you collect data in a format suitable for analytics
- How can you augment DevOps with AI/ML
 - E.g. predict time-to-merge of pull request

GitHub PR TTM: Approach



Code Repository

- Collect data using [ML tool](#)

```
python -m srcopsmetrics.cli \  
--create \  
--repository foo_repo \  
--entities PullRequest,Issue,Commit
```

- types of files, # lines, author, etc.

GitHub PR TTM: Approach



Code Repository



Feature exploration
& Engineering

- Transform features for ML
- Feature selection using mutual information, chi sq, etc

GitHub PR TTM: Approach



Code Repository



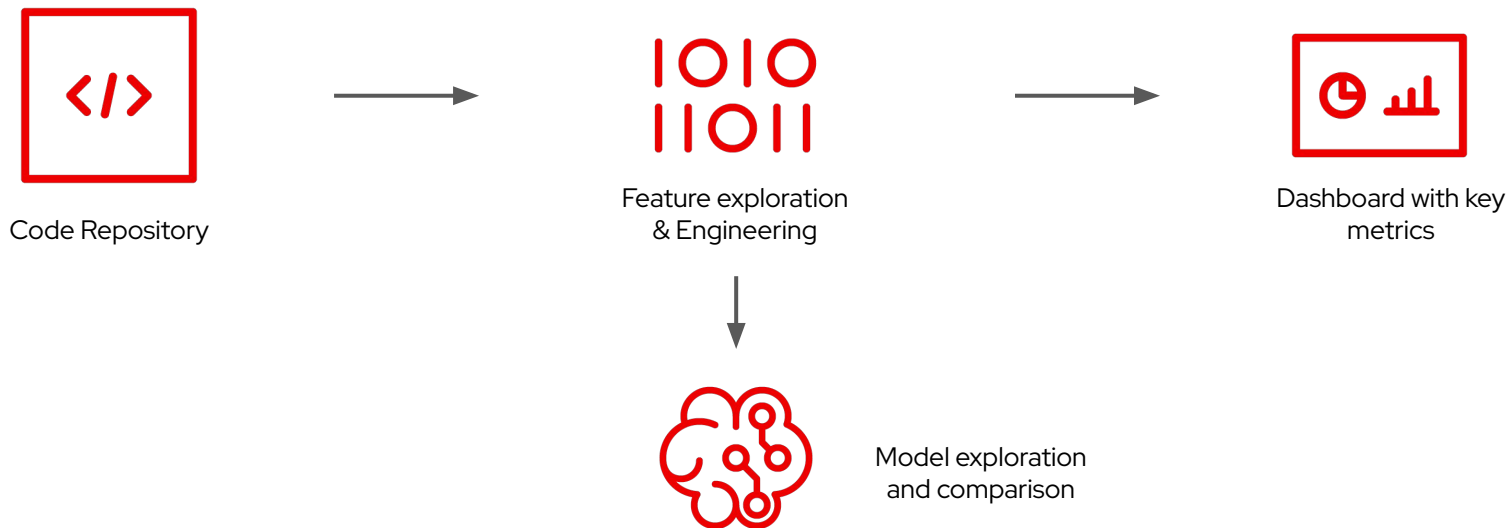
Feature exploration
& Engineering



Dashboard with key
metrics

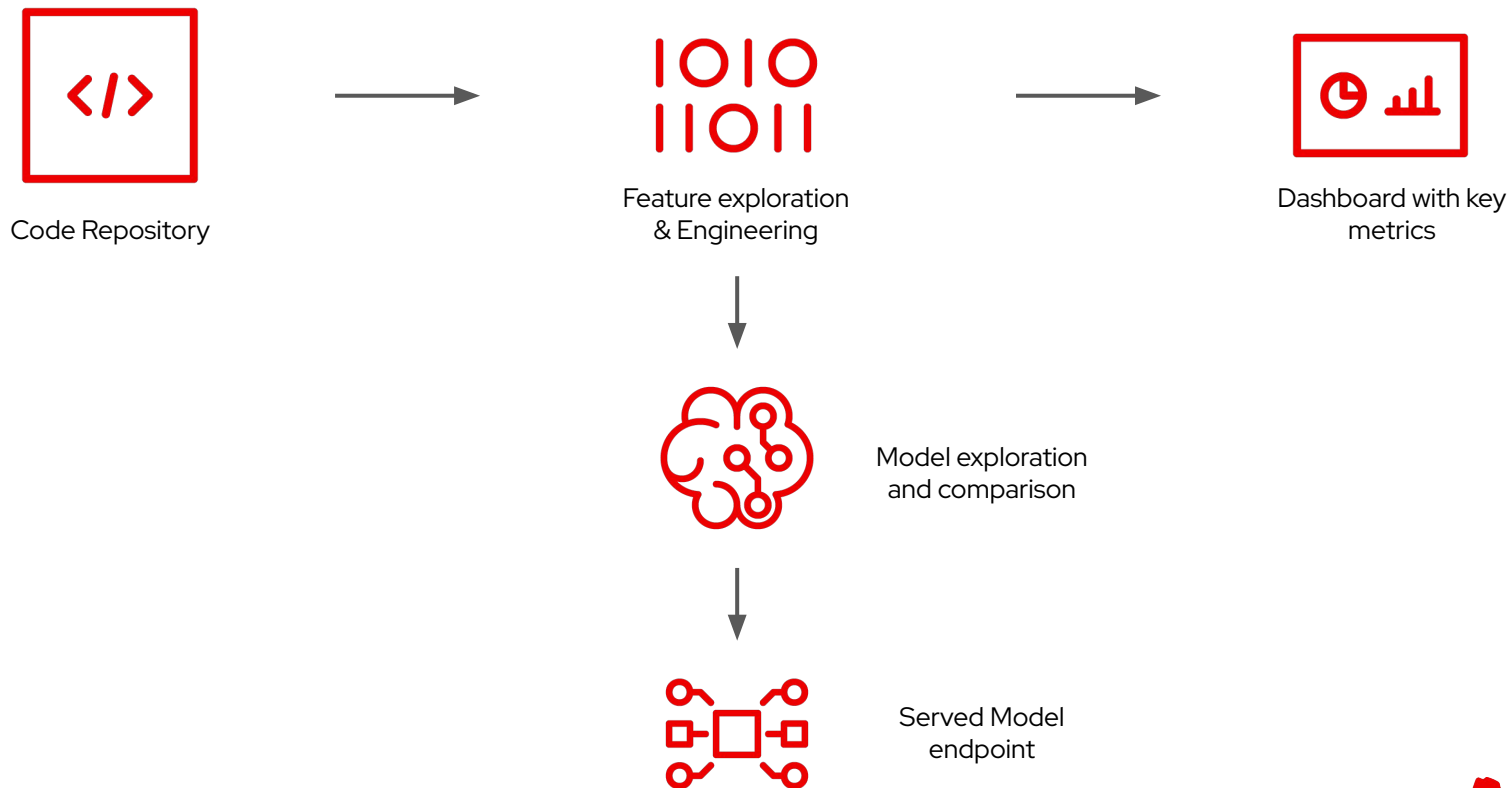
[Superset dashboard](#)

GitHub PR TTM: Approach



- Classification and regression setups
- SVM, RF, XGB, etc.

GitHub PR TTM: Approach



GitHub PR TTM: Service Endpoint

- Service deployed as a custom Seldon inference server
 - <http://github-pr-ttm-ds-ml-workflows-ws.apps.smaug.na.operate-first.cloud/predict>
- Demo notebook
 - https://github.com/aicoe-aiops/ocp-ci-analysis/blob/master/notebooks/time-to-merge-prediction/model_inference.ipynb

GitHub PR TTM: Next Steps

- Generalize model across repos
- Integrate into Thoth bots
- Set up model monitoring
- Reiterate on model

Overview

- What is AI4CI
- Tools and Services
- Machine Learning Models
 - Github Time to Merge
 - **Optimal Stopping Point**
 - Build Log Clustering
- Engage

Optimal Stopping Point: Overview

Problem to Address

Every new Pull Request to a repository with new code changes is subjected to an automated set of builds and tests before being merged. Some tests may run for longer durations than expected.

Longer running tests are often painful as they can block the CI/CD process for longer periods of time. How can we optimize the running time of our tests and prevent bottlenecks?

Optimal Stopping Point: Overview

Problem to Address

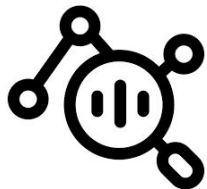
Every new Pull Request to a repository with new code changes is subjected to an automated set of builds and tests before being merged. Some tests may run for longer durations than expected.

Longer running tests are often painful as they can block the CI/CD process for longer periods of time. How can we optimize the running time of our tests and prevent bottlenecks?

Solution

By predicting the optimal stopping point for a test, we can better allocate development resources.

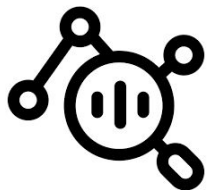
Solution Approach



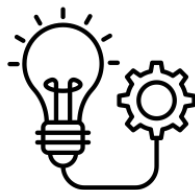
TestGrid Data

Test | Timestamps with Test Status

Solution Approach



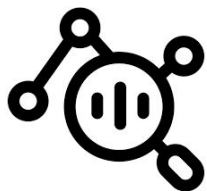
TestGrid Data



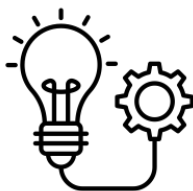
Feature Engineering

Find the **distribution type for passing and failing tests**. Probability density plots are used to find the probabilities of test duration

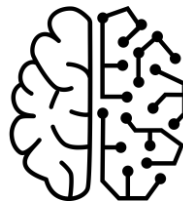
Solution Approach



TestGrid Data



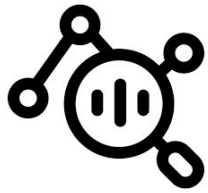
Feature Engineering



Model Training

Predict optimal stopping point by finding the point where:
probability of failure > probability of passing

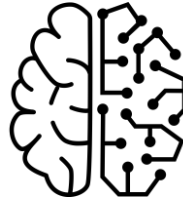
Solution Approach



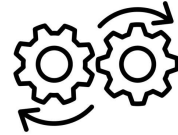
TestGrid Data



Feature Engineering



Model Training



Model Deployment



Model Interactions

- Feature Engineering notebook:
https://www.operate-first.cloud/data-science/ai4ci/notebooks/data-sources/TestGrid/metrics/probability_to_fail.ipynb
- Model training notebook:
https://www.operate-first.cloud/data-science/ai4ci/notebooks/optimal-stopping-point/osp_model.ipynb
- Model service deployed as a custom Seldon inference server:
<http://optimal-stopping-point-ds-ml-workflows-ws.apps.smaug.na.operate-first.cloud/predict>
- Model inference notebook:
https://www.operate-first.cloud/data-science/ai4ci/notebooks/optimal-stopping-point/model_inference.ipynb

Overview

- What is AI4CI
- Tools and Services
- Machine Learning Models
 - Github Time to Merge
 - Optimal Stopping Point
 - Build Log Clustering
- Engage

Build Log Clustering

- Introduction
 - Openshift CI build logs
 - [Pep's work](#) on EDA and clustering
- [Endpoint](#)
- [Deployment config](#)
- [Demo notebook](#)
- Next steps
 - Create a model for build logs from other jobs

Overview

- **What is AI4CI**
- **Tools and Services**
- **Machine Learning Models**
 - **Github Time to Merge**
 - **Optimal Stopping Point**
 - **Build Log Clustering**
- **Engage**

Engage

Get Started

<https://www.operate-first.cloud/data-science/ai4ci/docs/get-started.md>

Open Data
Sources



Notebooks



Dashboards



Model Endpoints



Automated
Workflows



Video Playlist



Engage

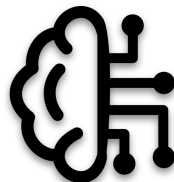
Contribute

<https://www.operate-first.cloud/data-science/ai4ci/docs/get-started.md>

KPI Metric



ML Analysis



Thank you, questions?



<https://github.com/aicoe-aiops/ocp-ci-analysis>



[@OperateFirst](#)



<https://tinyurl.com/aiforci>



<https://www.operate-first.cloud/data-science/ai4ci/>