

Automated Provision of Data Science Tools to End-Users using Docker & Kubernetes

Methods of data science, especially those dealing with machine learning are irreplaceable tools in many fields of science today. Without these methods, many experimental or observational results could not be transferred to useful scientific results. However, in many cases the researchers developing data science methods are not the end-users of them. Especially for today's complex models, such as in the field of deep-learning, highly specialized researchers are necessary to develop and implement appropriate methods for a specific analysis task.

This necessitates a quick and easy way to deliver those methods to the domain experts that perform the collection of data and have the necessary knowledge to interpret the results obtained by utilizing data science. Moreover, it poses the challenge that often times these domain experts are not experts in programming or interacting with non-GUI interfaces to run programs, which means that the data science researchers need to provide user-friendly access to their tools.

To solve these problems we have developed workflows combining tools for web-based GUI provision and tools for automatic provisioning of these tools to the researchers with none or minimal need for system administrator actions. Using Flask / Django and Shiny it is possible to easily create responsive, web-based user-friendly GUIs as a frontend for access to new algorithms or models tailored to a specific task. Using a Docker based development workflow, data science researchers can use git templates to integrate their algorithms into these GUIs. Using these templates, the resulting applications are automatically built using continuous integration / continuous delivery (CI/CD) pipelines and deployed to a Kubernetes based cluster. These workflows include testing and building and packaging the container, deploying the container to a Kubernetes Cluster using Rancher and ArgoCD as well as setting up appropriate SSL secured networking to the application for easy and secure web-access.

Together, this workflow allows to deploy finished data science applications to the end user in a few minutes and facilitates rapid updating and changing of data science methods to adjust to a specific task.

Primary authors: HNIOPEK, Julian (Friedrich-Schiller-University Jena & Leibniz Institute of Photonic Technology Jena); Dr STEFANIUK, Nazar (Friedrich-Schiller-University Jena & Leibniz Institute of Photonic Technology Jena); BOCKLITZ, Thomas (Leibniz Institute of Photonic Technology, Member of Leibniz Health Technologies, Member of the Leibniz Centre for Photonics in Infection Research (LPI))

Session Classification: Poster session