Red Hat

Top 5 ways developers and data scientists can collaborate

Data scientists don't bite

You need to know how data scientists work and how to work with them to build effective artificial intelligence (AI)-based applications. That means knowing the basics of AI and how to collaborate with your data science colleagues. Here are the top 5 things you need to know when working with data scientists and building AI-driven intelligent applications. Use this checklist as a guide to forming good working relationships and exceptional application development collaborations.

1 Understand how data scientists work

Data scientists are usually more concerned with building and refining their models than they are with application development or integrating their models into a piece of software. They rarely want to be involved in building continuous integration/continuous delivery (CI/CD) pipelines or writing application code and use tools you may not be familiar with, like Python, R, and Jupyter Notebooks.

They probably won't be the first ones to suggest an open collaboration with you–even though you are all working toward the same goals. However, solid, consistent teamwork between you and your data science team is essential to building good applications. Active collaboration results in:

▶ The deployment of intelligent, data-driven applications that effectively take advantage of AI.

▶ The opportunity for data scientists to put their modeling work to use in deployable solutions that add value to your company and its customers.

It will likely fall upon you to make the initial outreach and facilitate the collaborative experience with your data science colleagues. Adopt the guidance in this checklist to find out how to connect with your data science team in a beneficial way.

2 Find common ground

Explaining your development practices and seeing how they complement your data scientists' efforts is important to creating frictionless collaboration and an experience that works for everyone. To that end:

▶ **Encourage frequent touchpoints.** Setting up frequent and regular touchpoints is best to help ensure that the projects you work on together remain on track.

▶ **Respect boundaries.** Data scientists may not want or need to know how you get applications into production. Although MLOps is a popular concept, some scientists prefer to email you their Jupyter Notebooks. Respect their interests and the ways they like to work, and they will reciprocate.

▶ **Share each other's processes.** Besides learning how data scientists work, share your processes and the tools you use in production, like Git, Tekton, or Kubernetes. In the spirit of open source, give them a peek into your processes.

▶ **Use a common platform for collaboration.** Common cloud-native AI development platforms like Red Hat® OpenShift® AI support and encourage collaboration between you and your data science team. The platform democratizes the use of AI tools and allows teams to implement and accelerate intelligent application development.

3 Learn to work with model training tools

Learn at least the basics around some of the model training tools data scientists use regularly. Having working knowledge of model training tools will help you understand how the models are built. These are some of the most popular model training tools and libraries:

▸ Jupyter and PyCharm development environments

▸ PyTorch

▸ scikit-learn

▸ TensorFlow

Familiarizing yourself with these and other tools will improve your odds of successfully creating deployed model applications. It will also give you a better understanding of the work that goes into the creation of the models and help you to work out issues when models don't integrate into your intelligent applications smoothly.

4 Keep using your favorite tools and processes

When working with data scientists and AI, you are going to need to learn a lot of new processes and a few new tools. But you can keep using many of your favorite tools to do your logic. Application code and modeling can be done in any preferred language or framework.

For example, as a Quarkus developer you can do your application logic in Quarkus and have it make an application programming interface (API) call to a representational state transfer (REST) endpoint, while your data scientists handle the actual data processing and predictions—through a tool like Python or R. Although AI is complex, keep your work simple by using familiar tools and processes.

5 Remember the model is a part of the application

The model is important; so are the MLOps behind the model. Here are four things you will need to do to ensure your models continue to perform well when they are put into production:

▸ Build a model serving infrastructure that works for the application you're developing.

▸ Create new or extend existing CI/CD pipelines to handle both training and serving the model.

▸ Scale the model-serving application.

▸ Integrate streaming data services such as Apache Kafka and other data gathering components.

Finally, the deployment of your application is only the beginning. Models keep changing and will need to be monitored. Work closely with your data science team to define which metrics you or your operations counterparts need to monitor to prevent model drift. If an issue or change occurs, collaborate with your data scientists to refine and improve your models.

Learn more

Visit Red Hat OpenShift AI and read our ebook.

Give it a try

Try Red Hat OpenShift AI and preview a demonstration.

About Red Hat

Red Hat helps customers standardize across environments, develop cloud-native applications, and integrate, automate, secure, and manage complex environments with award-winning support, training, and consulting services.

f facebook.com/redhatinc
🐦 @RedHat
in linkedin.com/company/red-hat

North America
1 888 REDHAT1
www.redhat.com

Europe, Middle East, and Africa
00800 7334 2835
europe@redhat.com

Asia Pacific
+65 6490 4200
apac@redhat.com

Latin America
+54 11 4329 7300
info-latam@redhat.com