Docker Research

CPSC 491 Team 4

Sebastian Vargas

Docker is a platform for creating, developing and hosting containers and using containers is one way to package software to run on an isolated operating systems. In contrast to VMs, containers contain only the libraries, settings, and dependencies to make whatever software it hosts, work. This makes them efficient, lightweight and ensures that programs will run the same way regardless of what machine it is run in.

Source: https://www.docker.com/what-docker

Why use Docker to implement machine learning models?

Taken from this site: https://indico.io/blog/data-science-deployments-docker/

- No inconsistencies between team environment configurations:

Software configuration is always a pain. Docker’s configure once, run anywhere model means your teammates will have to worry less about environment setup and can focus more on writing code and building machine learning models.

- Reliable deployments:

Fewer bugs crop up in production when you can be assured that your development environment is identical to your production environment.

- Git-like tool for environment configuration:

If something does go wrong in production, reverting to a previous Docker image ensures you can quickly get back to a functional state.

Essentially, it let's developers (us) focus more on the machine learning models and less on having a reproducible and portable environment.

Also, using containers to run machine learning models makes them easier to run on a cloud since we do not need to worry about system incompatibilities.