Machine Learning Experiments Tracking

Mostafa S. Ibrahim Teaching, Training and Coaching for more than a decade!

Artificial Intelligence & Computer Vision Researcher PhD from Simon Fraser University - Canada Bachelor / MSc from Cairo University - Egypt Ex-(Software Engineer / ICPC World Finalist)



© 2023 All rights reserved.

Please do not reproduce or redistribute this work without permission from the author

Experiments Tracking

- Experiments tracking is the practice of logging, monitoring, and comparing different ML experiments
 - One major challenge is getting lost on what produced what!
- Reproducibility: ability to reproduce the models with the same settings and data

Experiments Tracking: Key Components to Track

- Code Version: Git hash / branch name
 - Model Architecture:
 - Use names and versions to track
 - You may copy significant files to the experiment folder
- Data Versioning
 - Keeping track of the version of the dataset used for training and validation.
- Configuration & Hyperparameters: E.g. LR, batch size, epochs
 - Save in files in your project. Copy to the experiment folder
- Metrics:
 - Save a json file with the different evaluation information and metrics results
- Artifacts: Output files like trained models, plots, or log files.
- Logs and Notes

Experiments Tracking: Tools for Experiment Tracking

- TensorBoard: Used mainly for TensorFlow projects, but also for PyTorch
- **MLflow**: An open-source platform that manages the ML lifecycle, including experimentation, reproducibility, and deployment.
- **Comet.ml**: Provides a cloud-based service for tracking, comparing, explaining, and optimizing experiments and models.
- Custom Solutions: Big companies typically have their in-house tools
- Use Google spreadsheets to summarize the different experiments and their files location
- Use Google docs to analyze the experiments and include samples with errors

"Acquire knowledge and impart it to the people."

"Seek knowledge from the Cradle to the Grave."