Survey on Logging Practices in Machine Learning-Based Applications.

The goal of this study is to look into logging practices in machine learning based applications. The survey seeks insights from practitioners to identify current practices, including where and what ML practitioners log inside their applications, and what ML logging libraries practitioners use. Your response to this survey will help advance the understanding and development of logging practices in the domain of machine learning.

Title of research project : Studying Logging Practice in Machine Learning-based Applications

Research Team:

Full Name: Foalem patrick loic

Position: PhD student

Polytechnique Montréal – Department of computer engineering and software engineering

Email: patrick-loic.foalem@polymtl.ca

Research project Supervisor

Full name: Foutse Khomh

Position: professor

Polytechnique Montréal – Department of computer engineering and software engineering

Telephone: 1-514-340-4711 ext. 4233

Email: foutse.khomh@polymtl.ca

Co-supervisor

Full name: Heng Li

Position: professor

University or faculty affiliation: Polytechnique Montréal

Do you accept to participate in this survey ?

*

Yes

No

Après la section 1

Rubrique 2 sur 10

Section 1: Demographics

Description (facultative)

Organization/Institution:

Position/Role:

*

Years of experience in software engineering:

*

1.

Less than 1 year

2.

1

3.

2

4.

3

5.

4

6.

5

7.

6+

Years of experience in Python programming:

*

1.

Less than 1 year

2.

1

3.

2

4.

3

5.

4

6.

5

7.

6+

Years of experience in machine learning:

*

1.

Less than 1 year

2.

1

3.

2

4.

3

5.

4

6.

5

7.

6+

Après la section 2

Rubrique 3 sur 10

Section 2: Current Logging Practices

Description (facultative)

Have you used logging in your machine learning-based application?

*

Yes

No

If yes, which components of the application do you typically log? (Select all that apply)

Data collection

Data preprocessing

Model training

Model evaluation

Inference/prediction

Autre…

What logging frameworks or tools do you commonly use in your machine learning projects? (List all)

Après la section 3

Rubrique 4 sur 10

Purpose(s) of using logging library

Description (facultative)

Have you used the default Python logging library for logging purposes in machine learning based application (e.g, : logging)?

*

Yes

No

If yes, please specify some context in which you use the default Python logging library.

Après la section 4

Rubrique 5 sur 10

Logging Information

Description (facultative)

What specific information do you log during Data collection stages

What specific information do you log during Data preprocessing stages

What specific information do you log during Model training stages

What specific information do you log during Model evaluation stages

What specific information do you log during Inference/prediction stages

Après la section 5

Rubrique 6 sur 10

Data management

Do you Agree or Disagree with the following statement ?

I find it beneficial to use logging for state events during data collection, such as recording data or processing erroneous data.

Agree

Disagree

I typically log data format information during data collection to keep track of the types or formats of the collected data.

Agree

Disagree

I log data information during data loading, such as file names and data formats, to ensure data integrity and traceability.

Agree

Disagree

In my machine learning projects, I employ data parallelism to accelerate data processing and training, and I log resource information like GPU/CPU usage and memory consumption during training.

Agree

Disagree

When I work on online learning ML systems, I log data schema to maintain consistency between different data streams.

Agree

Disagree

In the data processing phase, I encounter and log missing information to handle missing values, labels, and statistics about missing values.

Agree

Disagree

I find it helpful to log statistical information about datasets, including local estimators like mean, median, minimum, maximum, and variable estimators such as variance and standard deviation.

Agree

Disagree

Après la section 6

Rubrique 7 sur 10

Model management

Do you Agree or Disagree with the following statement ?

I agree that model management is a critical part of machine learning development, involving the development, evaluation, validation, and release of models into production.

Agree

Disagree

In my experience, I use logging to track experiments and generalize on new data, allowing me to reproduce or compare past experiments and select the best model.

Agree

Disagree

In my experience, I use logging to track and manage experiments performed during the model management phase.

Agree

Disagree

I log relevant information during model parallelism, such as state events, memory consumption, and GPU/CPU information about the devices used.

Agree

Disagree

During the model training phase, I use logging to record information about time, evaluation of training time, or time taken to complete an epoch.

Agree

Disagree

I have logged data about memory consumption, hyperparameters, and various metrics (e.g., accuracy, precision, f1-score) during model training.

Agree

Disagree

I employ state events logs to track the beginning, errors, hyperparameter updates, and end of the training process during model training.

Agree

Disagree

I have used logging to record information about the best model, memory consumption, and performance measures (e.g., accuracy, precision, f1-score) during the model evaluation phase.

Agree

Disagree

In my machine learning projects, I log GPU/CPU device information during model prediction and record time latency to monitor inference execution time.

Agree

Disagree

I am aware of the significance of logging optimizer information during model deployment, especially when different optimizers are used for real-time inferences.

Agree

Disagree

Après la section 7

Rubrique 8 sur 10

Configuration management

Do you Agree or Disagree with the following statement ?

I have used logging to record information about important libraries used for dependencies configuration to ensure the ML module of my application works correctly.

Agree

Disagree

I have logged information related to the environment settings, such as library versions and CUDA devices, to ensure the proper functioning of my machine learning application.

Agree

Disagree

Après la section 8

Rubrique 9 sur 10

General Comments

Description (facultative)

Do you have any suggestions on improving the current ML logging practices or developing new tools for ML logging?

Après la section 9

Rubrique 10 sur 10

Acknowledgement

Description (facultative)

Acknowledgement

Thank you for participating in this survey! Your input is greatly appreciated in advancing our understanding of best practices and challenges in logging for machine learning-based applications.