Introduction

You are receiving this survey as our initial investigation identified you as either a contributor to a machine learning project or as a professional software developer. This survey should take approximately 10 minutes to complete.

With this survey, we want to know your opinion regarding practices pertaining to the maintenance of machine learning software.

Your participation is voluntary and confidential. You can withdraw at any time. We do not record any identifying information.

This survey is conducted by a team of computer science researchers from University of California, Irvine (UCI).

The results of this survey will be used in a scientific publication. Should you be interested in being informed about the outcome of this study or any resulting publication, please contact Jiri Gesi (fjiriges@uci.edu).

By clicking the button below, you acknowledge that your participation in the study is voluntary, you are 18 years of age, and that you are aware that you may choose to terminate your participation in the study at any time and for any reason.

Please note that this survey will be best displayed on a laptop or desktop computer. Some features may be less compatible for use on a mobile device.

- ○ I consent, begin the study
- ○ Maybe later

Demography Questions

Who do you currently work for?

- ○ Company or Organization
- ○ University (including student)

- Independent/Freelancer/Consultant
- Other

What best describes your primary work area?

- Data & Applied Science. Example job titles: Applied Scientist, Data Scientist, Data Scientist Lead, Applied Science Manager, etc.
- Program Management. Example job titles: Program Manager, Program Manager Lead, etc.
- Research. Example job titles: Researcher, Lead Researcher, Research Manager, etc.
- Software Engineering. Example job titles: Software Engineer, Software Development Engeneer, Software Engineering Manager, etc.
- Other ____ Please enter an 'other' value for this selection.

In what location do you work?

- North America: USA
- North America: Canada, Mexico
- Central America and South America
- Europe
- Asia: China
- Asia: India
- Asia: Middle East
- Asia: Japan
- Asia: South Korea
- Asia: Other
- Australia, New Zealand, Oceania
- Africa
- Other ____ Please enter an 'other' value for this selection.

How many people do you work with on a regular basis on machine learning projects (approximately)?

- No one (I don't need to coordinate my work)
- 1-2 people

- 3-5 people
- 6-14 people
- 15-29 people
- 30 or more people

How many machine learning projects do you work on right now?

This question requires a valid number format.

How many years of work experience do you have?

This question requires a valid number format.

How many years of work experience do you have with machine learning related projects?

This question requires a valid number format.

How many years has your current team been working with machine learning?

This question requires a valid number format.

Where have you acquired your core machine learning skills and knowledge? Select all that apply.

- [ ] Formal education (e.g., BSc, MSc, PhD).
- [ ] Internal training programs.
- [ ] On the job (e.g., by building AI products).
- [ ] Online materials (e.g., Coursera, Edx, Udacity, YouTube).
- [ ] Self-taught.
- [ ] Other ____ Please enter an 'other' value for this selection.

Verify low Level Smells

Below are some machine learning project development challenges. Have you ever seen similar situations during machine learning project development? If yes, please select to what extent the situation impacts your project maintenance. If no, please continue to the next situation.

	Yes	N
1. Have you ever worked on a machine learning project in which you could use some available third party machine learning library/frameworks, but your team programmed it by yourselves instead?	○	○
2. Have you ever worked on a machine learning project in which you should have used specific library functions, but your team used some more generic library functions instead that introduced more parameters to define?	○	○
3. Have you ever worked on a machine learning project in which your team defined all functions with public access, and later discovered conflicts, because the same function name is used in other classes/files, resulting in having to change some of these public functions to private?	○	○
4. Have you ever worked on a machine learning project in which your team tested the machine learning models directly in the model training files instead of creating specific new functions to test machine learning models?	○	○
5. Have you ever worked on a machine learning project where many newly defined functions lacked default parameters, which made the functions error prone and difficult to understand?	○	○
6. Have you ever worked on a machine learning project that did not properly separate machine learning model training, testing, and data processing code into corresponding files?	○	○
7. Have you ever worked on a machine learning project where the machine learning models were programmed based on the developers' own understanding and interpretation but should strictly follow the original design of the model, which made the code difficult to understand?	○	○
8. Have you ever worked on a machine learning project that installed unused libraries or dependencies, which made the project difficult to deploy?	○	○
9. Have you ever worked on a machine learning project that hard-coded some literals (ex: path, model name) in strings, rather than using variables?	○	○
10. Have you ever worked on a machine learning project that failed to remove unwanted debugging statements?	○	○
11. Have you ever worked on a machine learning project that imported the same package/module multiple times (ex: using "import A" and "from A import B" at the same time)?	○	○
12. Have you ever worked on a machine learning project that used some third party machine learning libraries/frameworks (ex; tensorflow, pytorch), where updates/changes made in third party libraries/frameworks raised problems in the project and forced developers to make changes?	○	○

Besides the above code maintenance challenges in machine learning software, have you encountered other challenges while maintaining machine learning software systems? Please explain.

Verify old code smells

Below are some high level AI project development scenarios. Have you ever seen a similar scenario during AI project development? If yes, please choose how much these scenarios made your project's maintenance more difficult. If no, please go to the next scenario.

	See	
	Yes	No
1. Have you ever worked on a machine learning system in which a change to one component (ex: a feature, hyper-parameter, etc.) required making changes to multiple other components?	○	○
2. Have you ever worked with a machine learning system that involved running multiple distinct models sequentially (i.e. the output of one model was used as the input for the next model)?	○	○
3. Have you ever worked on a machine learning system in which results from your models may have been implemented for external parties to use in their systems but did not inform you?	○	○
4. Have ever worked on a machine learning system in which input features were unstable, i.e. they would change over time?	○	○
5. Have you ever worked on a machine learning system that contained a feature that became redundant over time but may not have been immediately identified as redundant?	○	○
6. Have you ever worked on a machine learning system in which there was a push to improve model accuracy even though the accuracy gain would have been marginal or there would have been high complexity overhead?	○	○
7. Have you ever worked on a machine learning system where the implementation of a supporting package within a code hindered the ability to improve or modify the model?	○	○
8. Have you worked on a machine learning system in which the data pipeline appeared to have grown with development needs over time as opposed to being thoughtfully planned out beforehand?	○	○
9. Have you ever worked on a machine learning system that incorporated bundled components in which not all components were actually needed?	○	○
10. Have you ever worked on a machine learning system that contained obsolete code paths that may have been originally added for experimental or testing purposes?	○	○
11. Have you ever worked on a machine learning system that used multiple programming languages in a way that did not seem efficient?	○	○

Besides the above high level maintenance challenges in machine learning software, have you encountered any other challenge while maintaining machine learning software

system? Please explain.

Powered by Qualtrics