Adoption and Effects of Software Engineering Best Practices in Machine Learning

Alex Serban, Koen van der Blom, Holger Hoos, Joost Visser

Leiden University, Radboud University, Software Improvement Group
The Netherlands

Software engineering best practices for machine learning

- Adoption of **machine learning** technologies calls for mature **engineering** techniques

- We aim to empirically determine the **state of the art** in how teams develop, deploy and maintain software with ML components

- We mined both academic and grey **literature** and identified 29 engineering practices for ML applications

- We conducted a **survey** among 313 practitioners to determine the degree of adoption for these practices and to validate their perceived effects

Software Engineering for Machine Learning

Leiden Institute of Advanced Computer Science (LIACS), The Netherlands

An ever-increasing number of organisations are developing applications that involve machine learning (ML) components. The complexity and diversity of these applications calls for engineering techniques to ensure they are built in a **robust** and **future-proof** manner.

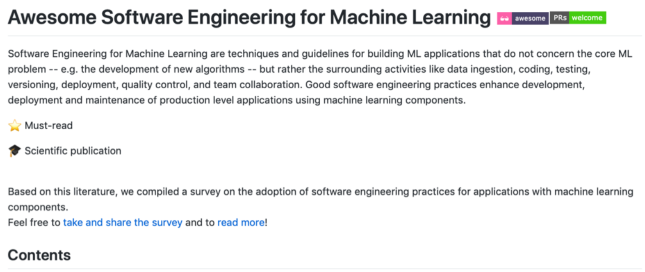
On this website we collect, validate and share engineering best practices for software including ML components. To this end, we study the scientific and popular literature and engage with machine learning practitioners.

For more information access our catalogue of ML engineering best practices or read our annual report on the State of Engineering Practices for Machine Learning.

https://se-ml.github.io

Literature review on SE for ML

- We reviewed over **50 articles**, both from academic publications and grey literature
- The majority of literature on this topic comes from **grey literature**
- We compiled a curated list of articles, available on GitHub, as an **awesome list**
- We compiled a **catalog** of 29 SE practices for ML, in 6 categories, corresponding to the ML development process


Awesome Software Engineering for Machine Learning awesome PRs welcome

Software Engineering for Machine Learning are techniques and guidelines for building ML applications that do not concern the core ML problem -- e.g. the development of new algorithms -- but rather the surrounding activities like data ingestion, coding, testing, versioning, deployment, quality control, and team collaboration. Good software engineering practices enhance development, deployment and maintenance of production level applications using machine learning components.

⭐ Must-read

🎓 Scientific publication

Based on this literature, we compiled a survey on the adoption of software engineering practices for applications with machine learning components.
Feel free to take and share the survey and to read more!

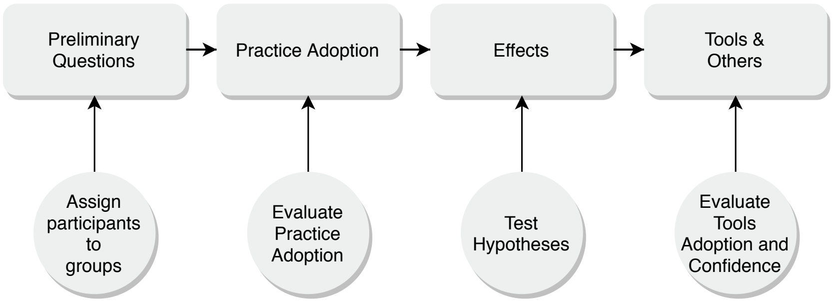
Contents

https://github.com/SE-ML/awesome-seml

https://se-ml.github.io/practices

Study design

- To validate the practice adoption we designed an observational study with **45 questions** in 4 sections and 4 possible answers (Likert scale)
- The **preliminaries** allow participants to be assigned to groups (concurrent control study)
- The practice **adoption** section evaluates the relevance and adoption of each practice
- The **effects** questions allow to test the hypothesis that adopting a set of practices leads to an intended effect



Survey demographics and practice adoption

- Europe has a higher **presence**, although other regions are also well represented
- The **adoption** of practices is similar across regions, except North-America, where adoption is higher
- **Tech** companies lead in practice adoption
- Practice adoption increases with team **size** and **experience**

(a) Respondents grouped by regions.

(b) Respondents grouped by organisation type.

(c) Respondents grouped by team size.

(d) Respondents grouped by team experience.

(a) Adoption of practices grouped by regions.

(b) Adoption of practices grouped by organisation type.

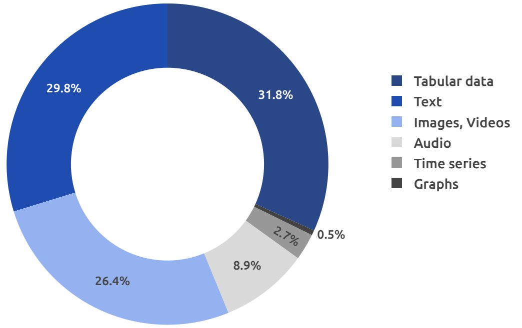
(c) Adoption of practices grouped by team size.

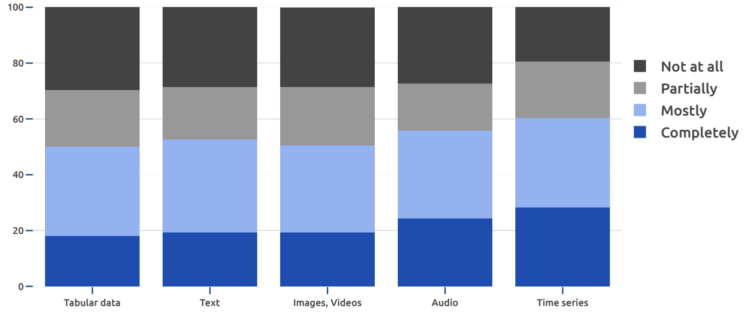
(d) Adoption of practices grouped by team experience.

Based on 313 valid answers

Data and practice adoption

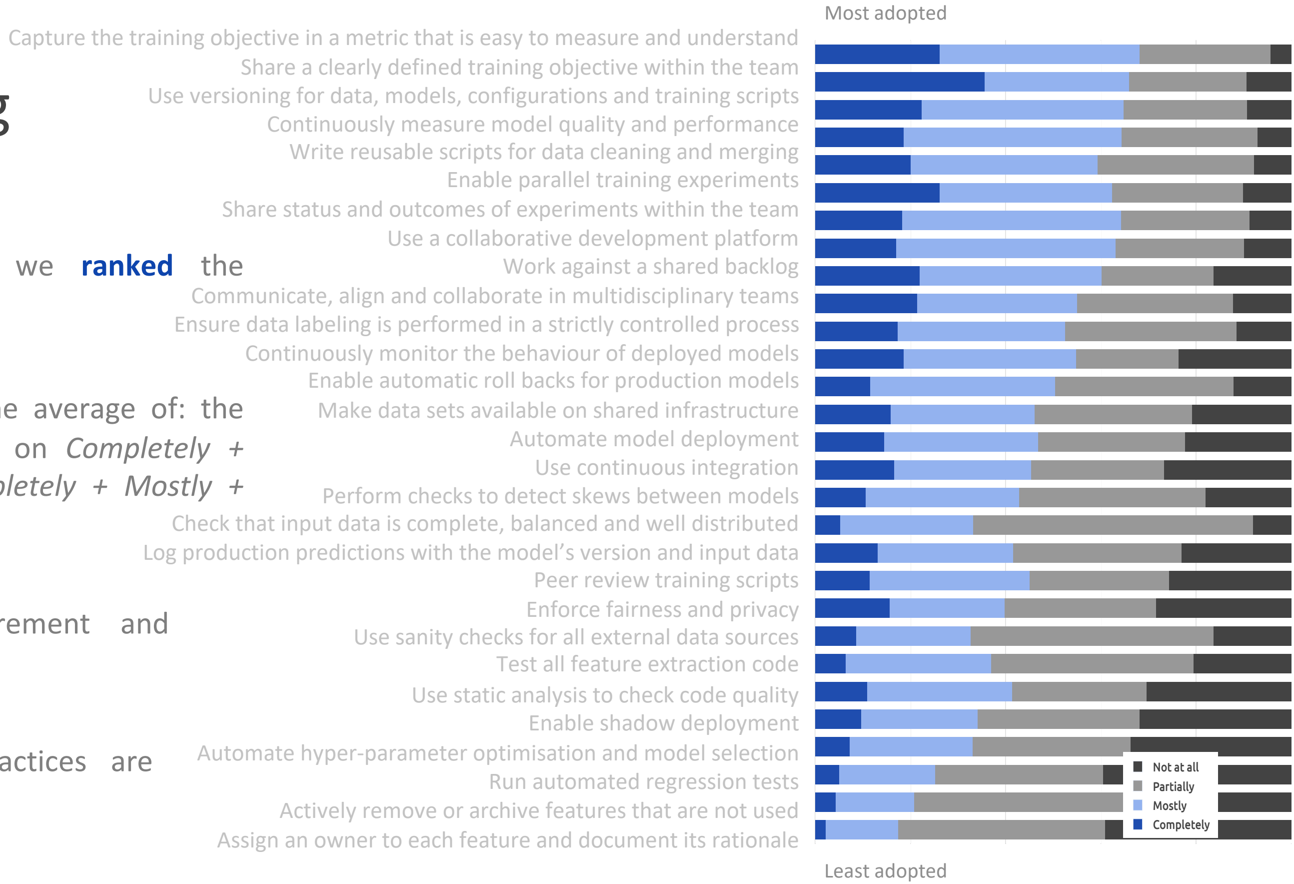
- The adoption of practices is largely **independent** of the data type used
- Teams that work with **Audio** and **Time Series** data show higher practice adoption
- Although these teams are also less represented
- We omitted Graphs data types due to lack of data





Practice ranking

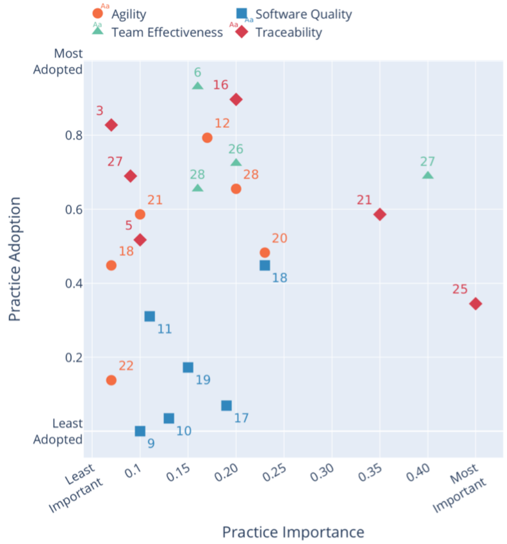
- Using the survey answered we **ranked** the practices
- The **ranking algorithm** uses the average of: the rank on *Completely*, the rank on *Completely + Mostly* and the rank on *Completely + Mostly + Partially*
- Practices related to measurement and versioning are **widely adopted**
- The two most **neglected** practices are related to feature management



Relationship between practices and effects

Effects	Practices	MSE / R^2 / ρ Linear Regression	MSE / R^2 / ρ RF	MSE / R^2 / ρ RF Grid Search	MSE / R^2 / ρ AutoML
Agility	12, 18, 20, 21, 22, 28	0.69 / 0.44 / 0.68	0.27 / 0.78 / 0.92	0.25 / 0.80 / 0.92	**0.24 / 0.82 / 0.92**
Software Quality	9, 10, 11, 17, 18, 19	0.35 / 0.71 / 0.83	**0.12 / 0.90 / 0.91**	0.17 / 0.87 / 0.91	0.17 / 0.87 / 0.91
Team Effectiveness	6, 26, 27, 28	0.45 / 0.63 / 0.87	0.25 / 0.80 / 0.90	0.19 / 0.84 / 0.92	**0.18 / 0.85 / 0.92**
Traceability	3, 5, 16, 21, 25, 27	0.38 / 0.69 / 0.80	0.22 / 0.82 / 0.90	**0.21 / 0.83 / 0.93**	0.22 / 0.82 / 0.93

- **4 effects** were built in the survey: Agility, Software Quality, Team Effectiveness and Traceability
- In all cases, the adoption of practices correlates strongly with the effects, and the effects can be **predicted** from practice adoption
- For each effect, we plot the practice **importance** as revealed by the predictive models, and their adoption rate
- The plot gives an overview of which practice to adopt first, in order to **maximize** return of investment for achieving a desired effect



Alex Serban
cs.ru.nl/~aserban

Conclusions

- We surveyed academic and grey literature and identified 29 software engineering **practices** for ML
- We validated practice **adoption** and their **effects** through a survey with 313 participants
- We built an open source **reading list** and a **catalogue** of engineering practices for ML
- We plan to **extend** the questionnaire with more practices, and create an **assessment** tool for teams developing ML apps
- **Contributions** to the reading list and the catalogue are welcomed!!

Reading list

Check out the awesome list with relevant literature: https://github.com/SE-ML/awesome-seml

Catalogue

Check out the catalogue of ML Engineering Practices: https://se-ml.github.io/practices

Survey

Take the survey yourself: https://se-ml.github.io/survey

Report

Check out the 2020 State of Engineering Practices for ML report: https://se-ml.github.io/report2020