Machine Learning Software Defect Prediction in vivo: Thirteen Considerations

Szymon Stradowski[1,2] Lech Madeyski[1]

38th IEEE/ACM International Conference on Automated Software Engineering (ASE 2023)

[1] Department of Applied Informatics
Wroclaw University of Science and Technology
Poland

[2] Nokia, Mobile Networks, Radio Frequency
Poland

Wrocław University of Science and Technology

NOKIA

Research area

Improving Product Quality and Decreasing Costs by Employing Machine Learning Software Defect Prediction

1
Exploring the challenges in software testing of the 5G system at Nokia: A survey (IST'23 [1])

2
Machine learning in software defect prediction:
A business-driven systematic mapping study (IST'23 [2])

3
Industrial applications of software defect prediction using machine learning:
A business-driven systematic literature review (IST'23 [3])

4
Can we Knapsack Software Defect Prediction?
Nokia 5G Case (ICSE'2023 [4])

5
Software Defect Prediction and its Industrial Application in Nokia 5G System-Level Testing (ongoing)

6
Bridging the Gap between Academia and Industry in Machine Learning Software Defect Prediction: Thirteen Considerations (ASE'23)

[1] Stradowski & Madeyski (2023). Exploring the challenges in software testing of the 5G system at Nokia: A survey. Information and Software Technology, 153:107067.

[2] Stradowski & Madeyski (2023). Machine learning in software defect prediction: A business-driven systematic mapping study. Information and Software Technology, 155:107128.

[3] Stradowski & Madeyski (2023). Industrial applications of software defect prediction using machine learning: A business-driven systematic literature review. Information and Software Technology, 159:107192.

[4] Stradowski & Madeyski (2023). Can we Knapsack Software Defect Prediction? Nokia 5G Case. In ICSE'2023 Companion., pp. 365-369.

Wrocław University of Science and Technology

NOKIA

Experience report

Our underlying research **aims to complement the existing system-level test practices with additional ML SDP mechanisms** within an extensive and complex software quality assurance process for cutting-edge 5G wireless communication technology development in Nokia [10,11].

Nokia 5G is a **real industry context** our observations and conclusions come from.

We provide a sequenced guideline containing thirteen consecutive considerations practitioners should account for when planning to introduce ML SDP in an industrial environment.

We built it upon the shoulders of the global standard of the business analysis body of knowledge [16].

[10] The 3rd Generation Partnership Project. 3GPP REL15. https://www.3gpp.org/specifications-tchnologies/releases/release-15

[11] Nokia Corporation. Nokia Annual Report 2021. https://www.nokia.com/system/files/2022-03/nokia-ar21-en.pdf

[16] IIBA, Babok: A Guide to the Business Analysis Body of Knowledge. International Institute of Business Analysis, 2015, https://www.iiba.org/career-resources/a-business-analysis-professionals-foundation-for-success/babok/

Wrocław University of Science and Technology

NOKIA

1) Collect requirements and set appropriate goals.

Gathering requirements and setting proper goals is critical to the success of any business endeavor [16].

Preceding the ML SDP introduction, a widespread survey was launched among test practitioners within the company to elicit opinions on the current challenges [1].

The following high-level goals have been defined:

- The proposed solution must not disrupt the already existing quality assurance processes.
- Defect prediction efficiency must be on an acceptable level (not necessarily super high, but good enough).
- Built framework should utilize existing data and should be fully automated.
- There are no immediate timeline requirements for the project.
- Cost-effectiveness is important (positive return on investment (ROI)).
- Prediction modelling needs to allow interpretability

[16] IIBA, Babok: A Guide to the Business Analysis Body of Knowledge. International Institute of Business Analysis, 2015

[1] Stradowski & Madeyski (2023). Exploring the challenges in software testing of the 5G system at Nokia: A survey. Information and Software Technology, 153:107067.

Wrocław University of Science and Technology

NOKIA

2) Build upon solid theoretical and practical foundations.

Theoretical preparation before starting a project is important for several reasons:

- it helps to define the scope and goals of the project correctly,
- helps to ensure that the team has a thorough understanding of the problem they are trying to solve,
- enables effective planning by providing the information needed to create a detailed project plan,
- and improves overall decision-making,
- it helps the research community to validate shared results [20].

For our theoretical preparation, we have used rapid reviews [21] to adapt the regular review process to fit the practical constraints.

[20] E. Barr, C. Bird, E. Hyatt, T. Menzies, and G. Robles, "On the shoulders of giants," in Proceedings of the FSE/SDP Workshop on Future of Software Engineering Research. ACM, 2010, pp. 23–28.

[21] B. Cartaxo, G. Pinto, and S. Soares, "The role of rapid reviews in supporting decision-making in software engineering practice," in EASE '18: 2018. New York, NY, USA: ACM, 2018, pp. 24–34.

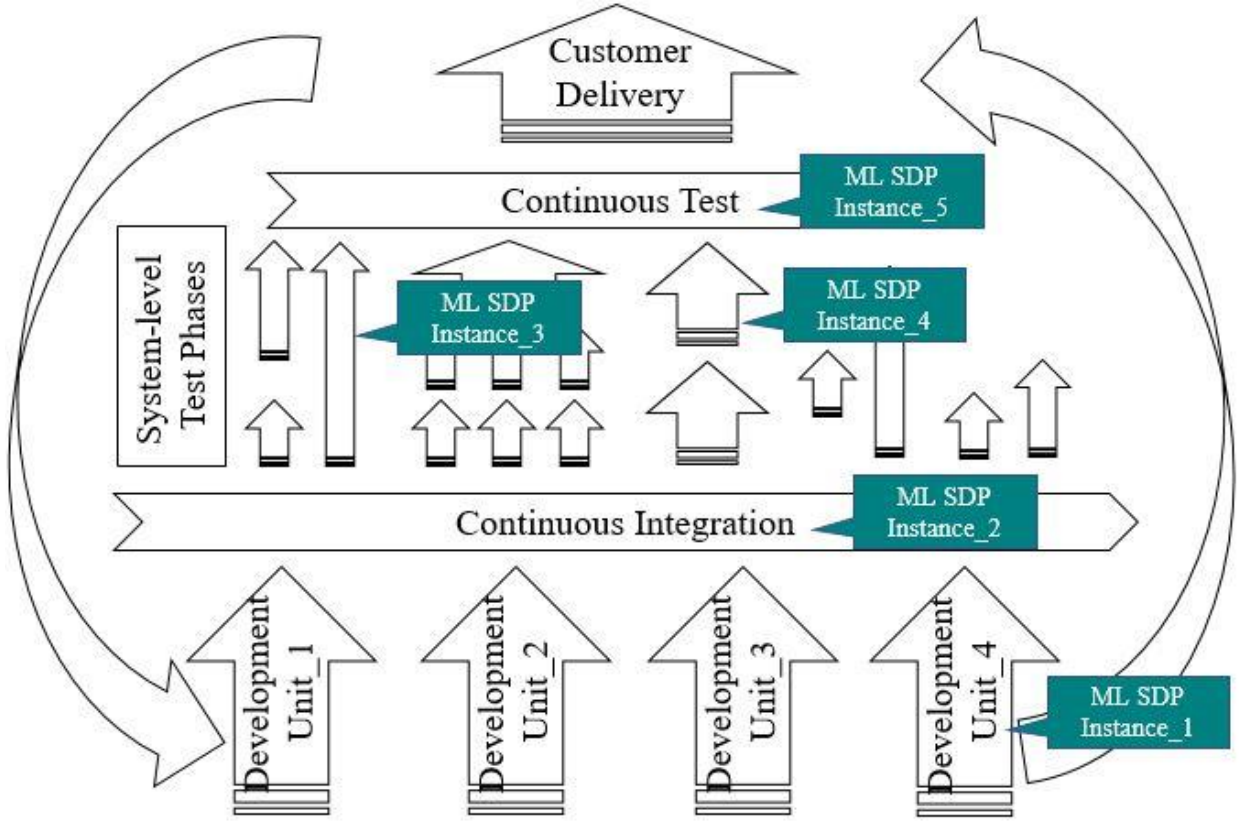
Wrocław University of Science and Technology

NOKIA

3) Consider the entire SDLC.

Large-scale software development requires scaling methodologies to manage efficiently [12], and testing with a single-layered verification effort is rarely possible for grand products.

A well planned and executed ML SDP introduction can extend to all of the test phases within the life cycle, providing even more substantial saving potential to the company [8].



[12] H. Edison, X. Wang, and K. Conboy, "Comparing methods for large-scale agile software development: A systematic literature review," IEEE Transactions on Software Engineering, vol. 48, no. 08, 2022.

[8] S. Stradowski and L. Madeyski, "Can we Knapsack Software Defect Prediction? Nokia 5G Case," in 2023 IEEE/ACM 45th International Conference on Software Engineering: Companion Proceedings (ICSE'23). New York: IEEE/ACM, 2023, pp. 365–369

Wrocław University of Science and Technology

NOKIA

4) Conduct technology assessment and introduction.

The goal of the technology readiness level assessment (TRL) is to optimize preparation for a new technology change within the company [32].

Noteworthy steps were the following:

- TRL 1: Basic principle observed - thorough theoretical preparation as described in Consideration 2).
- TRL 2: Technology concept formulated - consideration of the context and defining requirements and goals as in Consideration 1).
- TRL 5: Technology validated in a relevant environment - lightweight working solution to make initial inroads.
- TRL 6: Technology demonstrated in a relevant environment - a working pilot solution finalized by a showcase with main stakeholders.
- TRL 8: System complete and qualified - fully working and tested solution (currently ongoing).

[32] P. Raffaini and L. Manfredi, "Chapter 15 - project management," in Endorobotics, L. Manfredi, Ed. Academic Press, 2022, pp. 337–358

Technology Readiness Assessment Guide - U.S. Government https://www.gao.gov/assets/gao-16-410g.pdf

Wrocław University of Science and Technology

NOKIA

5) Conduct risk analysis.

Risk assessment is the process of identifying and evaluating potential hazards that might endanger the success of a project, as well as analyzing what mitigations can be triggered when the hazard occurs [16].

The impact and probability scores can be multiplied to calculate the risk value used to prioritize the effort being committed to response actions (we planned four response types to the risks: avoid, mitigate, accept, or transfer).

TABLE I
EXAMPLE RISK ANALYSIS.

Risk#	Risk name	Impact (3-1)	Probability (4-1)	Risk value	Response
Risk 1	Lack of available resources	3	4	12	Mitigate
Risk 2	Lack of needed competence	2	2	4	Avoid
Risk 3	No visible containment gain	3	1	3	Avoid
Risk 4	Tooling and license unavailability	3	1	3	Mitigate
Risk 5	Organizational changes	2	2	4	Accept
Risk 6	Resistance among practitioners	3	3	9	Transfer
Risk 7	Loss of critical data	3	1	3	Avoid

[16] IIBA, Babok: A Guide to the Business Analysis Body of Knowledge. International Institute of Business Analysis, 2015

Wrocław University of Science and Technology

NOKIA

6) Choose appropriate data set.

Commercial companies can possess vast amounts of data that can be used for ML SDP. Regarding the entire SDLC (as in Consideration 3)), predictors for each test phase can be built on very different data sets.

Consequently, the data gathered and utilized determine many aspects of the implementation:

- conducting feature selection on the data set is important to optimize the process and increase prediction performance,
- commercial data often suffer from missing samples, especially if the data set contains manually entered fields in the repository,
- many additional concepts are expanding the opportunities to use ML SDP in different circumstances that can be considered (JIT, CCDP, CPDP, HDP, etc.) [37-41].

[37] F. Hutter, L. Kotthoff, and J. Vanschoren, Automated machine learning: methods, systems, challenges. Springer Nature, 2019.

[38] Z. Zeng, Y. Zhang, H. Zhang, and L. Zhang, "Deep just-in-time defect prediction: How far are we?" in ISSTA 2021. ACM, 2021, p. 427–438.

[39] S. Hosseini, B. Turhan, and D. Gunarathna, "A systematic literature review and meta-analysis on cross project defect prediction," IEEE Transactions on Software Engineering, vol. 45, no. 2, pp. 111–147, 2019.

[40] Y. Ma, G. Luo, X. Zeng, and A. Chen, "Transfer learning for crosscompany software defect prediction," IST, vol. 54, no. 3, pp. 248–256, 2012.

[41] J. Nam and S. Kim, "Heterogeneous defect prediction," in Proceedings of the 2015 10th ESEC/FSE 2015, 2015, p. 508–519.

Wrocław University of Science and Technology

NOKIA

7) Choose appropriate tooling.

Choosing the framework to use for solution introduction in vivo needs to be based on the requirements and goals defined in Consideration 1).

Importantly, licensing and other legal aspects of the used framework and obtained results must be carefully verified.

Automated ML enables the process of gathering the data, pre-processing, simulation, and presenting the result to be acted upon to be fully automated [38].

Another essential factor to consider is the compatibility of the ML framework with existing data repositories within the company [3].

A dedicated solution can be built (insourced or outsourced). However, considering the high quality of already available frameworks, good reasons need to exist for building something new.

[38] F. Hutter, L. Kotthoff, and J. Vanschoren, Automated machine learning: methods, systems, challenges. Springer Nature, 2019.

[3] Stradowski & Madeyski (2023). Industrial applications of software defect prediction using machine learning: A business-driven systematic literature review. Information and Software Technology, 159:107192.

Wrocław University of Science and Technology

NOKIA

8) Apply appropriate learners and performance metrics.

Many ML techniques have been conceived each with different prediction effectiveness based on the circumstance (as in the "no free lunch" (NFL) theorems [45]). Therefore, we advise employing various classifiers to select the best-performing ones for used data sets

Also, additional techniques such as normalization, outlier detection, feature selection, re-sampling and cross-validation, hyperparameter tuning, boosting, ensuring reproducibility, and conducting statistical analyses need to be considered depending on the context. Before committing to highly effective but also very complex solutions such as deep learning, alignment with the solution requirements (Consideration 1)) is advised.

Significantly, selection should be based on reliable performance metrics. There are many performance measures; however, considering the arguments and recommendations [14,15], the main comparisons and conclusions should rely on Matthew's Correlation Coefficient (MCC).

[45] D. H. Wolpert and W. Macready, "No free lunch theorems for optimization," IEEE Transactions on Evolutionary Computation, vol. 1, no. 1, pp. 67–82, 1997.

[14] Yao and M. Shepperd, "The impact of using biased performance metrics on software defect prediction research," Information and Software Technology, vol. 139, p. 106664, 2021.

[15] D. Chicco and G. Jurman, "The matthews correlation coefficient (mcc) should replace the roc auc as the standard metric for assessing binary classification," BioData Mining, vol. 16, no. 1, p. 4, 2023.

Wrocław University of Science and Technology

NOKIA

9) Build for interpretability.

We advise the principle of limited trust towards any machine learning predictions, especially in business-critical contexts. Predictive technology can analyze vastly more extensive amounts of data than any human could, but it may also lack aspects like experience, ethics, or intuition [48,49].

Also, several reasons can influence decisions behind building-in interpretability [50] potential to the ML SDP solutions: Transparency and trust, Accountability and compliance, Debugging and improvement, Domain expertise, Knowledge discovery.

Notably, high interpretability (high model transparency) typically comes at the cost of performance. If a company wants to achieve the highest performance but still wants to explain the ML model's behavior in human terms, model explainability may be the way to choose.

[48] A. Kotriwala, et al, "Xai for operations in the process industry-applications, theses, and research directions," in AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering, 2021, pp. 1–12.

[49] L. von Rueden, et al, "Informed machine learning – a taxonomy and survey of integrating prior knowledge into learning systems," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 1, pp. 614–633, 2023.

[50] A. Barredo Arrieta, et al., "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Information Fusion, vol. 58, pp. 82–115, 2020.

Wrocław University of Science and Technology

NOKIA

10) Prepare a cost evaluation.

Cost-effectiveness is a critical aspect of any project from the business perspective (adhering to the value-based software engineering (VBSE) [53] concept).

Unfortunately, the cost and benefits of ML SDP are not yet well understood, and the scientific research on the topic is scarce [3, 43].

General cost model we used to present the business case of ML SDP introduction in Nokia was based on a publication by Herbold [43]. The proposed model includes aspects such as the initial investment needed to set up the solution, expenses related to running the simulations, additional quality assurance effort, and escaped defects. For our purpose, we calculated the ROI value, which showed a very positive outlook. Also, using popular economic ratios has helped to gain management support for the project.

[53] B. Boehm, "Value-based software engineering: Reinventing," SIGSOFT Software Engineering Notes, vol. 28, no. 2, p. 3, mar 2003.

[3] Stradowski & Madeyski (2023). Industrial applications of software defect prediction using machine learning: A business driven systematic literature review. Information and Software Technology, 159:107192.

[43] S. Herbold, "On the costs and profit of software defect prediction," IEEE Transactions on Software Engineering, vol. 47, pp. 2617–2631, 2019.

Wrocław University of Science and Technology

NOKIA

11) Manage stakeholders.

Stakeholder management is critical to any successful business endeavor, as stakeholders can considerably impact the project's outcomes [16].

- The first step we took in stakeholder management was identifying all individuals or groups with an interest or stake in the project.
- Second, we analyzed the stakeholder needs, interests, and expectations regarding the project (also in the context of requirements gathering, as in Consideration 1)).
- In consequence, a management strategy was developed based on assessing stakeholder requirements and priorities, containing plans for engaging and communicating with relevant parties throughout the project.
- Finally, a resource plan with team roles and responsibilities has been created and approved.

[16] IIBA, Babok: A Guide to the Business Analysis Body of Knowledge. International Institute of Business Analysis, 2015, https://www.iiba.org/career-resources/a-business-analysis-professionals-foundation-for-success/babok/

12) Plan for long-term evolution.

It is also worthwhile to plan for the long-term evolution of the ML SDP field in the coming years: evaluate possible scenarios and consider their feasibility in a specific context to seek further business opportunities [16].

In the future, will it be feasible to substitute human interference completely and entirely rely on ML SDP to make the best test coverage decisions?

Acknowledging the long-term future and feasibility of pure ML SDP solutions has heavily impacted the projects' technology assessment (Consideration 4)) and cost evaluation (Consideration 10)) steps.

TABLE II
STEP-WISE EVOLUTION OF ML SDP INTRODUCTION.

	Step 1: Current state Test Architects	**Step 2: Our solution TAs + ML SDP**	**Step 3: Future Pure ML SDP**
Predictive accuracy	baseline	improved	improved
Auto data acquisition	limited	yes	yes
Historical data	limited	yes	yes
Report generation	manual	automatic	automatic
Multiple projects	limited	limited	yes
Running time	full-time	low	low
Installation cost	high	low	moderate
Maintenance cost	high	low	moderate

[16] IIBA, Babok: A Guide to the Business Analysis Body of Knowledge. International Institute of Business Analysis, 2015, https://www.iiba.org/career-resources/a-business-analysis-professionals-foundation-for-success/babok/

Wrocław University of Science and Technology

NOKIA

13) Plan project closure.

At the end of a project, several key tasks should be planned to ensure a successful conclusion and handover [16].

- First of all, review the project goals and requirements and measure the achievements against the success criteria Consideration 1)).
- A handover of the responsibility towards designated practitioners needs to be done. It should be planned in advance, and the transfer of knowledge about ML and the used framework needs to be secured (Consideration 2)).
- Run a dedicated retrospective meeting to draw final conclusions on the process, gather lessons learned, and elicit stakeholder feedback (Consideration 11)).
- Reflect upon the next steps: conclude the technology acceptance model (Consideration 4)) and review the whole SDLC impact to identify further improvement opportunities (Consideration 3)).
- Finally, consider publishing the results and experience reports to benefit the wider community.

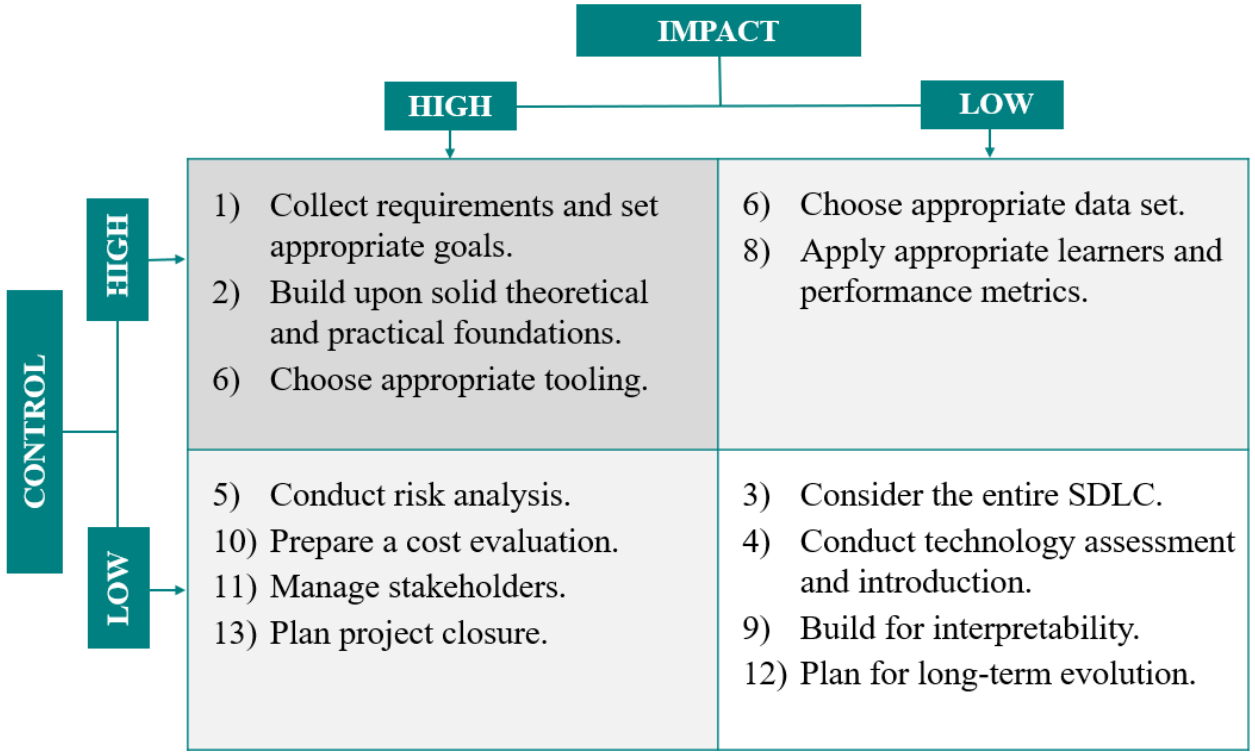
[16] IIBA, Babok: A Guide to the Business Analysis Body of Knowledge. International Institute of Business Analysis, 2015, https://www.iiba.org/career-resources/a-business-analysis-professionals-foundation-for-success/babok/

Wrocław University of Science and Technology

NOKIA

Practical importance

Control-Impact analysis:



Each consideration is evaluated on a scale from 1-low to 4-high:

- Impact - subjective measure reflecting how critical a particular consideration is to the end success of the ML SDP introduction.
- Control - subjective measure reflecting how much influence the project leaders can have over the outcome of the consideration.

Analysis provides which considerations influence the chances of final success at the lowest amount of time and effort spent.

Note: above considerations are an example reflecting the particular context of our research in Nokia. Potential followers must prepare their analysis according to the specific circumstances they are working in, as the results may be very different.

Wrocław University of Science and Technology

NOKIA

Thank You

Acknowledgment: This research was carried out in partnership with Nokia and was financed by the Polish Ministry of Education and Science 'Implementation Doctorate' program (ID: DWD/5/0178/2021).

Wrocław University of Science and Technology

NOKIA

Thirteen considerations

The proposed checklist consists of a sequence of thirteen steps provided below:

1) Collect requirements and set appropriate goals.
2) Build upon solid theoretical and practical foundations.
3) Consider the entire SDLC.
4) Conduct technology assessment and introduction.
5) Conduct risk analysis.
6) Choose appropriate data set.
7) Choose appropriate tooling.
8) Apply appropriate learners and performance metrics.
9) Build for interpretability.
10) Prepare a cost evaluation.
11) Manage stakeholders.
12) Plan for long-term evolution.
13) Plan project closure.

Wrocław University of Science and Technology

NOKIA