Participant Information Sheet

Title of Study: *Machine Learning Best Practices used by Software Engineering Researchers*

Department: Faculty of Computer Science and Mathematics, University of Passau

Name and Contact Details of the Researcher(s): Mojica-Hanke, Anamaria and Herbold, Steffen,

Faculty of Computer Science and Mathematics, University of Passau, Dr.-Hans-Kapfinger-Str. 30, 94032 Passau, Gemany,

- **Invitation Paragraph** You are being invited to take part in this research project. You should only participate if you want to; choosing not to take part will not disadvantage you in any way. Before you decide it is important for you to understand why the research is being done and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information.
Thank you for reading this.

- **What is the project's purpose?** The aim of this project is to get a better understanding of what software engineering researchers consider to be best practices when using machine learning within their research.

- **Why have I been offered to participate?** Because you are the first author of a paper at a premier software engineering venue of a paper that uses machine learning.

- **Do I have to take part?** It is up to you to decide whether or not to take part. You can withdraw at any time without giving a reason. If you decide to withdraw before completion, all your data will be discarded.

- **What will happen to me if I take part?** The study duration (including reading this information sheet) is less than 5 minutes. It takes place online. All data will be collected and stored in accordance with the General Data Protection Regulation.

- **What are the possible disadvantages and risks of taking part?** There are no foreseeable discomforts, disadvantages and risks for taking part.

- **What are the possible benefits of taking part?** Our research may aide your own future work through a better understanding of best practices when using machine learning.

- **What if something goes wrong?** If you would like to raise a complaint about this research, please contact the Principal Researchers Anamaria Mojica-Hanke and/or Steffen Herbold.

- **Will my taking part in this project be kept confidential?** All information that we collect is strictly anonymized and cannot be traced back to you. We do not collect any personal data. The organizers of this study of no means to identify any of the participants and you will not be identified in any ensuing reports of publications.

- **What will happen to the results of the research project?** We plan to disseminate the results in a publication in the ACM Transactions on Software Engineering and Methodology (TOSEM). You will not be identifiable in any report or publication.

- **Deception** There are no deceptive elements in this survey and you receive all information up-front.

- **Data Protection Privacy Notice** We will not process any of your personal data. The legal basis for contacting you in the first place is covered under the public interest exception of the General Data Protection Regulation (GDPR).
All data is anonymized.

- **Who is organizing and funding the research?** This study is organized by the research groups for AI Engineering headed by Prof. Dr. Steffen Herbold from the University of Passau, Germany in collaboration with Prof. Mario Linares-Vásquez from the Universidad de Los Andes, Colombia, and Prof. Denys Poshyvanyk from the College of William and Mary, USA. There is no third-party funding for this research.

Contact for further information Mojica-Hanke, Anamaria, Faculty of Computer Science and Mathematics, University of Passau, Dr.-Hans-Kapfinger-Str. 30, 94032 Passau, Gemany,

Thank you for reading this information sheet and for considering taking part in this research study.

*By clicking on **Begin** you agree to the terms and conditions outlined above for participation in this research study.*

Survey:

Question: What machine learning best practices have you used in your software engineering research papers, and how often did you apply those practices *(Always, Frequent, Sometimes, Never)*?
Instruction given bellow the question:
Please write each practice in a new line:
- Good practice [how often]