Teaching Predictive Modeling to Junior Software Engineers: Seminar Format and Its Evaluation

Katsiaryna Labunets
University of Trento
Trento, Italy
katsiaryna.labunets@unitn.it

Andrea Janes
Free University of Bozen-Bolzano
Bolzano, Italy
andrea.janes@unibz.it

Michael Felderer
University of Innsbruck
Innsbruck, Austria
michael.felderer@uibk.ac.at

Fabio Massacci
University of Trento
Trento, Italy
fabio.massacci@unitn.it

Research Context and Questions

"If you went to bed last night as an industrial company, you’re going to wake up today as a software and analytics company" (c) Jeff Immelt, CEO of General Electrics

– Machine learning techniques are widely used in research, but their usage is still limited in industry <= software developers lack for relevant knowledge.
– Software developers who work in company do not have time to attend long courses with lectures on prediction modeling.

Research Questions

RQ1: How could a seminar to teach prediction models be designed to promote the use of the method among novices?
RQ2: How effective is the chosen format of the seminar to teach prediction models?

Solution

Two-day seminar to teach machine learning-based predictions in software engineering. The seminar is based on the research paper by Scandariato et al. "Predicting vulnerable software components via text mining" [1].

Why this paper?

1) one of the authors knew this work and it was easier to replicate it,
2) security is a fascinating topics for the students => higher interest to participate in the seminar,
3) the paper was published in IEEE TSE journal => this is the high quality study.

Evaluation

RQ1: Two-day seminar with good balance between theoretical and practical components and based on a research paper as a scenario of the practical part.
Participants appreciated the practical illustration how machine learning techniques can be used to solve a problem that they would never think is possible to apply to.

RQ2: Participants demonstrated high overall quality at the level of Knowledge (78%), Comprehension (72%) and Application (68%), and medium quality of results at the Analysis level (59%).

Results

What to improve

1) The seminar should be at least two full days of work.
2) The seminar should provide a better explanation of "how things work" in prediction modeling as the participants experienced problem in this part.
3) The Analysis level requires better support by the seminar structure as participants showed fair quality of responses at this level. E.g. a more interactive format of the practical exercises can help participants to practice skills related to Application and Analysis levels.

References