

Review of a PhD Thesis

Introduction – the subject of the review

1. Legal basis for the review

- Act of 20 July 2018 – The Law on Higher Education and Science (Journal of Laws of 2018, item 1669 as amended);
- Resolution of the Academic Advancement Board of the Poznań University of Economics and Business of 8 April 2022 regarding appointing me as a reviewer of the doctoral dissertation (Letter of the Head of the Academic Advancement Board prof. Barbara Jankowska, PhD, of Poznań University of Economics and Business of 27 April 2022).
- I received from the Head of Academic Advancement Board of the Poznań University of Economics and Business via post official documentation about PhD dissertation of Mrs. Izabella Krzemińska, M.Sc. on 10 May 2022.

2. Evaluation of the relevance of the thesis.

The thesis deals with the very relevant topic of personalizing mobile phone apps based on an automatically computed user profile timely and on the client-side. The thesis adopts the Big 5 personality traits and predicts those using information solely available on the client device ensuring the privacy of user-sensitive data. The computation of the user profile is also done very fast (few seconds) and this makes the system react real-time. The two proposed applications UISPP, which builds the user profile, and PAPA, which uses the user profile for personalization serve as proof-of-concepts.

With the current explosion of mobile apps it is difficult to satisfy the needs of all users. That is why this thesis comes at the right time by offering personalization of mobile apps to user characteristics increasing the user satisfaction when using these. It also draws useful knowledge from psychology, machine learning, and software engineering, having a true multi-disciplinary character, a necessity when dealing with complex problems. The fact that the PhD student was embedded in several teams at Orange for experimenting with the proposed concepts is also a plus, showing the relevance of this work in industry.

Using a thorough literature, survey the thesis identifies the gaps in the current works, and proposes solutions on how to address these. At the end, existing solutions to user profiling are revisited and compared to the proposed ones. The thesis is relevant not only for academics but also for practitioners, e.g., making use of transpilers in porting Python code to Java code so that it can be deployed on mobile devices. Also, all the proposed features are realistic and the thesis shows how to gather these from the user device (with the necessary approvals by the user).

3. Evaluation of the content and structure of the thesis.

The thesis is well-structured and rich in content. It has a natural build-up relying on various theories to back up the carried processes. For example it uses a Design Science framework (DSR) to design the research methodology, a Literature Review framework (SLR) to survey the relevant literature (for profile building and personalization), and an Evaluation framework (FEDS) to evaluate the proposed methods. The thesis starts by investigating the literature, then it proposes UISPP for profile building and PAPA for an example application that makes use of UISPP, and, last, it evaluates the adequacy of UISPP (quantitatively and qualitatively) and PAPA (only qualitatively).

For the literature survey the thesis is also extensive as there are many personality modelling approaches considered, as well as machine learning solutions for building the profile, refuting the least useful ones. The information is nicely summarized in tables looking at the various characteristics of the considered approaches. The student properly considers the pros and cons of the methods found, and makes well-informed decisions. The considered features for machine learning are also explained in detail and the two proposed methods (UISPP and PAPA) are evaluated extensively (e.g., formally using the personality gathered information as well as informally based on users opinions).

There are not many studies that use psychology theory for personalization in software engineering projects, which also contributes to the novel character of this work. The personality questionnaire seems also carefully designed and analysed. Also, the use of linguistic information to be used for devising features for machine learning is well-explained. The application of the machine learning methods is properly done using cross-validation. Also, I appreciate that ensemble methods (including the variants of soft and hard voting) were also considered and these turned out to give the best results. In addition, the normalization and feature selection techniques seem to further boost the results.

The student has a good command of English, there is only a small number of typos. The sentence structure is elaborate with a careful plan for argumentation. The same can be said about the chapters which present first the most important concepts before diving deep into the details. The presentation is also fostered by many tables and figures ensuring a pleasant read. The less important matter is structured in the appendixes which are rich in details.

4. Evaluation of the merits of the dissertation (include originality, novelty and contribution to the discipline).

The proposed methods seem original and probably the highest level of novelty can be found in the considered features for determining the user profiles. These are extensive and carefully designed for the considered task. The random forest with the selected features leads to the best results. With an accuracy of around 70%, this goes beyond what is commonly found in the literature with only the eye-tracking experiment giving better results but this at the expense of rather slow computer vision computations which make it more difficult to have the application react in real-time.

The thesis also contributes to the literature with extensive surveys on user profiling and personalization. Also, there seem to be no current applications like PAPA, which makes a one-to-one comparison with the state-of-the-art difficult. The research takes also a modern approach to personalization accounting for user privacy (data is stored only on the client machine) and real-time constraints. The work goes beyond a data science project with a clear software engineering component. Also automatic tools to bridge these two worlds have been considered.

The student showed a lot of creativity in devising features that are useful for user profiling. This is a considerable contribution to the field of personalization. While personalization is not new, how to do it solely on the client is new. The user profiling part looks solid, the personalization is more of a sketch and needs more work. Nevertheless, as the main focus of the thesis is on user profiling, a non-trivial process, this is considered enough for the novel contribution of this thesis.

5. Concluding remarks

This is a carefully written thesis with a clear novel contribution to the field of mobile app personalization. As the role of mobile apps in our life is expected to increase, how to make these apps attractive to end-users is an open question and this thesis makes a step in answering this question by showing how to build a user profile for personalization obeying all the current regulations. The student has a good command of English and shows that she can express herself at the highest academic level, testimony being also her previous peer-reviewed publications on the thesis topic. It would be nice to get answers to some of the following questions during the defense:

Q1: On which dataset are the hyperparameters optimized? What function is used for optimization: loss, accuracy, precision, recall, F1?

Q2: How is the algorithm importance defined in soft voting?

Q3: How are the categorical features encoded?

Q4: How are the personality labels (LOW, MEDIUM, HIGH) computed based on the questionnaire (the questionnaire has 25 questions, with 5 questions per dimension). Is the questionnaire designed by you and made publicly available?

Q5: It is claimed that the 5 dimensions of personality are orthogonal, what does it mean to be an extravert and not open at the same time?

Q6: Can the SOLID technology pushed by Sir Tim Berbers-Lee be used for implementing the proposed solution?

Q7: How are the profiles updated after initialization? How often and with what information?

Q8: How is the personalization defined for PAPA?

Q9: Why are the neural networks not part of the considered machine learning approaches? These have shown very good results for many different tasks.

Q10: How can agreeableness be negatively correlated with a leader success?

I hereby declare that the doctoral dissertation entitled Automatic Data-Based Personality Assessment as a Method of Electronic Services Auto-Personalisation written by Mrs. Izabella Krzemińska, M.Sc. meets the requirements for doctoral dissertations according to the paragraph 187.2 of the Act of 20 July 2018 – The Law on Higher Education and Science (Journal of Laws of 2018, item 1669 as amended) and I recommend that the Academic Advancement Board of the Poznań University of Economics and Business approves Mrs. Izabella Krzemińska, M.Sc. for public defence of the dissertation.

In case a Reviewer requests the distinction of the dissertation please respect the following regulations of the Resolution No 57 (2020/2021) of the Senate of the Poznań University of Economics of 26 February 2021 on the procedure for awarding the degree of doctor at the Poznań University of Economics and Business:

§10.2 A request for awarding dissertation with a distinction may be submitted before or after the defence of the dissertation; however, at least two motions for awarding the distinction must be presented in the body of the review. A reviewer who has submitted a motion for awarding the distinction in the body of a review may withdraw their motion after the dissertation defence.

§10.3. A prerequisite for the procedure of recognising the dissertation as outstanding is submission of a motion for recognising the dissertation as outstanding by all reviewers.

Rotterdam, 1 July 2022

Flavius FrasinCAR

