IDEAS 2020 Track proposal on Software Engineering for Data Science

Juan J. Cuadrado-Gallego and Olga Ormandjieva.

Motivation for the proposal

In recent years the interest of the research and industrial communities about Data Science and its related knowledge fields, for example, Big Data or Internet of Things, has exceeded all expectations. The reason for this is the very interesting and useful applications that the Data Science projects have for the improvement of important aspects of the human life, such as effective communication, increased productivity, etc.

Data Science challenges have been profusely researched and improved, as the discipline of Data Science is deepening its roots in the industry and society, as probed, for example by the recently published book *The Data Science Framework (A view from the EDISON project)*, Springer International Publishing, 2020.

Perhaps one of the aspects that has delayed its advancement in relation with the other issues of the Data Science is the Software Engineering techniques and tools for developing Data Science projects. This a main issue that must not be overlooked, because of the enormous number of projects that are currently being developed and the importance that quality assurance has for these projects.

Data Science projects have their own characteristics, in many cases they are very different than other software development projects and new software engineering techniques and tools must be researched and developed in order to obtain the desired high-level quality for them.

This special track in the ACM conference IDEAS aims to help researchers share their ideas and obtain feedback as well as new ideas from their colleagues, dedicated to the new emerging field of software engineering for data science projects. In consequence this Special Issue invites contributions on the following topics (but is not limited to them):

- Software process for Data Science projects
- Data Science Engineering: principles and policies
- Data Science Quality in post-pandemic era: challenges and open problems
- Data Science and Big Data quality measurement frameworks
- Data Science system architectures: patterns and anti-patterns
- Managing Big Data waste: approaches and tool support
- Data Science and Big Data ethics
- Environmental sustainability of Data Science initiatives