# Introduction to the HICSS-54 Collaboration for Data Science Minitrack

Lina Zhou
Business Information Systems and Operations Management
The University of North Carolina at Charlotte
Charlotte, NC 28223
lzhou8@uncc.edu

Souren Paul
Graduate School of Computer & Information Sciences
Nova Southeastern University
3301 College Avenue
Fort Lauderdale, FL 33314
Souren.paul@gmail.com

Florian Schwade
University of Koblenz-Landau
Faculty of Computer Science
Institute for IS Research
Universitaetsstrasse 1
D-56070 Koblenz
fschwade@uni-koblenz.de

As the data size and diversity continuously grow, model and workflow complexities increase, and unstructured decisions soar, there is an increasing need for data science teams to collaborate for improved decision-making. Such collaboration could take place between various stakeholders both within and across different stages of data science projects. Collaboration in turn helps address data science challenges. Collaboration enables data scientists to be more productive and efficient in identifying relevant questions or problems, collecting data from different sources, organizing and making sense of the data, and communicating their findings in ways that can be easily used to support collaborative goals such as innovation, idea generation, decision making, negotiation, and problem solving. Therefore, collaboration is a key success factor in data science projects.

There has been a continued interest in looking for ways to increase value of data science and use it to address business challenges. Encouraging and facilitating collaboration in data science is one promising way for businesses and organizations to enhance their operational efficiency and/or competitive advantages. For example, human knowledge and experience can provide guidance in building computational models or in search of effective solutions to business problems.

This minitrack includes one paper session, consisting of three papers covering the following areas of interest: collaborative data science, collaborative analysis of big data, inter-organizational collaboration in data science, collaborative collection, aggregation, and organization of big data, human-guided knowledge discovery, and best practice for collaboration in data science.

The first paper, "Collaboration for Big Data Analytics: Investigating the (Troubled) Relationship between Data Science Experts and Functional Managers", examines the relationship between managers and data science experts as a potential reason for failure in big data analytics projects. Based on the social capital theory, it identifies that the relationship is troubled due to the absence of structural network ties and inconstant understanding of BDA applications in the business context.

The second paper, "Integrating Blockchain for Data Sharing and Collaboration Support in Scientific Ecosystem Platform", proposes Blockflow — a blockchain-based architecture to address the issues related to reliability, reproducibility, transparency, and privacy in collaborative research. It supports sharing provenance data in a trusted collaborative environment. The solution is evaluated in a feasibility study and the integration into a scientific software ecosystem is demonstrated.

The third and the last paper, "Unsupervised Deep Learning for Fake Content Detection in Social Media", detects fake social media content by developing unsupervised deep learning models. The models aim to address humans' cognitive bias in labelling training data and to improve efficiency in collaborative data science. By learning feature representations from the context of social media content, the proposed demonstrates in empirical evaluations superior performance to alternative unsupervised models and comparable performance to supervised models in fake content detection in social media.

URI: https://hdl.handle.net/10125/70640
978-0-9981331-4-0