Towards Reliable Data-Driven Computations

Abstract:

Data-driven methods are increasingly being used in domains such as fraud and risk detection, where data-driven algorithmic decision making may affect human life. The growing impact of data and data-driven systems on society makes it important that people be able to trust analytical results obtained from data-driven computations.

This can be done in two complementary ways: by providing result explanations so that the user understands the computation and the basis for the observed results; and by profiling and monitoring the data used in the computation, to make the results more reliable in the first place.

In the first part of the talk, I will present the use of provenance -- information regarding the data origin and computational process -- for providing explanations of computational results. In the second part of the talk, I will present a method for data profiling using labels, as an example of a data-focused technique to facilitate an analyst building a reliable decision-making pipeline.