This factsheet is a result of the first Wageningen University & Research (WUR) Data Science and AI Fellowship program. With this program we aim to increase and integrate our expertise in DS/AI throughout the entire organisation. The variety of projects highlights the potential for DS/AI across the WUR domains.

Author: Robert Smith

Life Science Data Integration Challenge

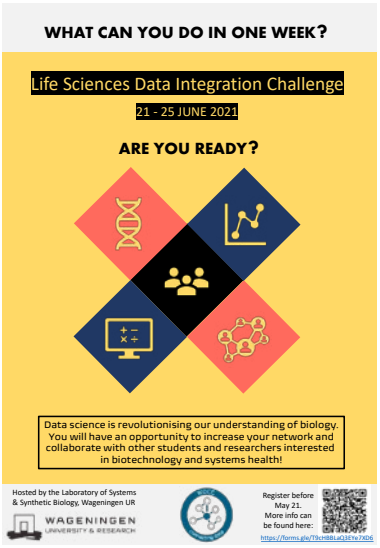
Building a bridge between mathematical models and data science for young researchers

Objective

We aimed to create a multi-disciplinary week-long challenge event where bioinformaticians and systems biologists combined skills to analyse data from the life sciences.

Method

There were four strands to challenge creation. First, we sourced large datasets and analysis tools from collaborators. Second, we advertised the challenge across relevant groups. Third, we organised a series of seminars to stimulate participants during the week. Finally, we created an online environment for participant teams to communicate and collaborate.



Results

The challenge hosted three teams of four participants. The participants educational background spread from bioinformatics, systems biology, and statistics to molecular biology. Due to the challenge being held online and widespread advertisement, participants joined from the Netherlands, Belgium, Italy, Greece, and Iran. In their teams, participants combined mathematical modelling, statistical methods, and machine learning methods to investigate anthocyanin regulation in tomatoes, links between people's oral health and microbial communities, and fatty acid production by microbes. Teams uncovered novel results from their analysis that will be followed up by researchers in the future.

Impact

Our challenge allows early career researchers to explore and combine methods taught during their data science educational programmes. The reviews we obtained from participants were largely positive, however they highlighted that the challenge should be restricted to those with the required background knowledge. Despite this concern, the winning team proposed a methodology that utilised mathematical models and automated data analysis pipelines to predict tissue-dependent expression of tomato enzymes. Such ideas of combining models and machine learning are gaining ground in literature, e.g. systems biology-informed neural networks*, showing the need to find innovational educational methods for students to explore these concepts.

Future plans

First, we have informed the data owners of the challenge team's results. It is likely that the novel results obtained during the challenge will lead to future research. Second, given the success of this initial challenge, we aim to run the challenge week again next year. In the next edition we shall increase awareness and advertisement such that we can obtain an increased number of bioinformatician and systems biology participants.

Further information

- *Przedborski et al. (2020) doi: 10.1038/s42003-021-02393-7
- Contact person: Robert Smith, Laboratory of Systems & Synthetic Biology
- Project duration: January – July 2021 with the challenge held June 21- 25.

WAGENINGEN
UNIVERSITY & RESEARCH