**Tilburg University takes up Data Challenge for Social Good**

A report of the Data Challenge kickoff

**Tilburg University is investing heavily in data science with new education programs, participation in the Graduate School for Data Science in Den Bosch, research, and building a data science ecosystem with companies and organizations in the Noord-Brabant region. But Tilburg University aims to go one step further, true to its social mission: it uses its data science knowledge and skills for the social good. In September, the first Data Challenge for Social Good was taken up.**

The idea was to solve data science problems brought in by non-commercial parties within a period of two weeks by students and academics. The number of problems that can be solved with help of big data analysis is growing: consider for instance the use of social media data to manage disaster response in case of an earthquake, which is the expertise of Kenny Meesters at the Tilburg School of Economics and Management. At the Data Challenge for Social Good event he made the point, however, that using big data can be extremely useful but may also conceal the fact that they represent real people and real problems that have to be solved. Making use of (big) data requires a shift in organizational thinking, he explained.

Another big data development that is already happening is the amount of health data becoming available through apps that measure health and lifestyle, personal health records being stored in ‘the cloud’, and health data shared by patients of particular diseases, for instance on the platform www.kanker.nl. As Tilburg University alumnus, e-health consultant and CEO of Xenios Bartho Hengst explained, the amount of data is growing steadily pushed by new technologies, but many questions still have to be solved. Who owns these health data and who has access to them? How can they be shared efficiently and securely by different stakeholders? Who can make real sense of them?

**Big Data, Big Challenges**

The reality of big data also proved more challenging than expected in the Data Challenge for Social Good, said Arjan van den Born in his introductory remarks. In front of a full house, he explained the hurdles of this (and any) data challenge.

First, the data have to be accessible and ‘clean’, so that they can be used in statistical and other analyses. Many NGO’s that had applied for the data challenge, sent data that were not ready to be analyzed.

Second, the issue at hand as well as the data set have to be manageable in size. If processing and analyzing data takes months or years, researchers need more time than a two week Data Challenge event that was foreseen initially. Most of the data challenges handed in did not meet those requirements.

But the Tilburg data scientists did take up two challenges that looked very promising. The results can however not be delivered within two weeks but everyone remains to committed to deliver them.

**“Data science may change doctors’ perspectives”**

One of the challenges taken up is that of Jeroen Bosch Hospital in Den Bosch brought in by pediatrician and Tilburg University professor Esther de Vries. She collected over 1.5 million health data by asking parents to report monthly on the health of a total of 761 basically healthy children between 2 and 18 year old over a period of two years. The aim is to determine what is normal as regards the frequency of having colds, headaches, flus, and the like. If we know what is normal, abnormal and hence potentially serious health issues such as immune disorders can be identified more easily.

Several researchers of the Methodology department at the School of Social and Behavioral Sciences led by prof. Marcel van Assen started to analyze the data and concluded that relatively simple statistical analyses are not going to give the answer. More complex analyses are still being carried out and possible other methods will have to be applied, such as machine learning and neural networks. In any case, data science may change doctors’ perspectives, said Esther de Vries, who was encouraged by the preliminary results.

**“Domain knowledge is crucial”**

The data challenge taken up was brought in by Greenpeace to determine which ships are involved in illegal fishing (an estimated 20% of all fishing activity), based on publicly available ship tracking data in the Automatic Identification System (AIS). Using machine learning and data mining, a research team led by Prof. Eric Postma of the Tilburg School of Humanities is going to detect anomalies in these data. The team is still working on extracting the right data to work with and coding them for analysis, which will actually be 80 percent of the work, but they are confident that some questions can be solved this way. Close cooperation with Greenpeace is pivotal though, explained Postma, since they have the domain knowledge that is crucial to select and analyze the data. With the results, Greenpeace should be able to alert authorities to illegal fishing anywhere in the world.

Concluding from the panel discussion that data science remains merely a tool to help finding answers, that it should begin as well as end with the people concerned and that one should not be too optimistic, Arjan van den Born promised that Tilburg University is committed to bring this Data Challenge for Social Good to a successful end. Next year, the second Data Challenge for Social Good will be bigger and better still.