MSc thesis internship – Machine Learning for Data Privacy

**Topic description**
Investigation into generative-adversarial learning for data privacy

ING views cooperation with the research community as an important element in staying amongst the most innovative players in the banking domain. Simultaneously, the scientific community has started various data-publishing initiatives in response to raised concerns about reproducibility. Banks will refrain from publishing their customers’ data, which is a problem for full participation in the scientific community. In addition to sharing data with the scientific community, sharing sufficiently realistic data might be valuable in collaborations with ‘ING externals’ such as contractors or consultants. Guarantees regarding privacy are a prerequisite for banks to participate in publishing and/or sharing any data. Other domains (e.g. healthcare) face similar challenges.

A possible long-term solution is to generate new data that is sufficiently similar to actual data to run experiments with. The data should be useable in some sense (e.g. to train a Machine Learning algorithm on) but should be sufficiently different from actual data to avoid de-anonymization.

The problem is to design, implement and test a robust framework for synthesising anonymous data for training Machine Learning. Generative Adversarial Networks (also known as Turing Learning) might provide a means to achieve privacy by design due to learning on synthetic rather than real data.

Questions to be considered include:
- How can we generate data that preserves privacy but is sufficiently realistic to train a model on?
- How can the quality of generated data be determined? Is this model-independent?
- How does using generated data instead of real data impact model performance?
- Are all privacy concerns covered by using generated rather than real data?

**About ING**
ING is transforming rapidly towards a top notch IT company with a Top Engineering culture, incorporating FinTech start-up mentality and financial innovation. We are end-to-end agile, work in multidisciplinary teams and use cutting edge technology. This enables us to offer our customers the best possible service and respond to rapidly changing customer demands.

You will be working in the ING’s Core Banking University, this is where science and business meet. Within this department top experts from both areas are collaborating on the next-generation digital bank.

**Requirements**
We are looking for a Master student that can meet the following requirements:

- Bachelor degree in a data science-related field, with great results. Interested in privacy concerns unsupervised learning, deep learning, generative models, etc.
- The ability to learn quickly and explore new topics
- Affinity with a cooperate research environment
- Data science experience (e.g. projects or working experience besides your study)
Fluent in English

**Interested?**
Please get in contact with us: send your CV and grade lists (BSc and MSc so far). If you have any other topic ideas, you can send them as well. Please be aware that a Certificate of Good Conduct and positive Pre-Employment screening are required.

Joost Bosman, IT Manager, ING: joost.bosman@ing.nl