Data science in Kela: cases and challenges

Aaro Viertiö
Product Owner – Data Science

Kela|Fpa

Data Science in Kela

- Data Science service since 4/2019
- Currently 5 data scientist
- Internal service, with all of Kela's functions as customers
- Agile practices often used in IT
- Focus on statistical and business aspects of data science with machine learning
  - Other teams focus on robotics, image recognition, language processing etc.

Case: Predicting social exclusion of young adults

- Social exclusion of youth is hot topic in Finland
- Huge costs to society, Kela is an important actor
- Kela started offering new multi-disciplinary service to high risk groups in 2018 to better serve those in the most difficult situations
- **Problem**: Could we identify who are at highest risk of social exclusion to direct them at our new service?

Case: Predicting social exclusion of young adults

- Trained a classification model to predict the expected level of response index in next 6 months with test data
  - Index includes many factors, for example probability of last resort benefits, homelessness, zero income
- Tuned to be accurate for the highest risk population at the expense of lower total accuracy
  - Over 98% accuracy for top 500 at risk

Case: Predicting social exclusion of young adults

- The hard parts:
  - Data often not available/reliable, eg. how to define a homeless person?
  - At which point is intervention most effective? Are we predicting too early/too late?
  - Is this possible with current legislation?

- The easy parts:
  - Building a good model
  - Coding

Case: Optimal length of social assistance decision

- The number of applications for social assistance (toimeentulotuki) has rapidly increased in past two years
- The decision is often made for only one month, even when the need for benefit is longer
  - Kela has started giving out longer (2, 6, 12 month) decisions to make benefit handling more efficient, but longer decisions often need to be revised later
- **Problem**: Can we predict for each application, how long should the decision be, to minimize handling costs taking into account the needed revisions?

Case: Optimal length of social assistance decision

- Trained a model to predict the chance of revision if a decision is made for X months
- Result of above model used as input for cost optimization algorithm
- The prediction and suggested length of decision could be served to the benefits officer who makes a decision

- Current model is very good and in testing, and we are building the production environment and front end. Hopefully in production Q2/2020.

Case: Optimal length of social assistance decision

- The hard parts:
  - Getting the application data from legacy system to data warehouse
  - Performance matters (often it doesn't!): with 200 000 applications per month, the models must serve over 20 per minute on average. Spikes of tens per second.

- The easy parts:
  - Building a good model
  - Problem very well defined

Real world data science

- Often used rule of thumb: data science is 20/10/2% analysis, and rest is cleaning data, trying to understand what the real problem is etc.
- What are the big challenges then?
  - Laws
  - Ethics
  - Data availability/quality and legacy systems
  - Getting the results to real use

The challenges

- Laws
  - GDPR, getting consent, automated decision making, cloud usage for identifiable info
- Ethics
  - We often work with those most vulnerable. Are the models robust for different groups?
- Data availability/quality and legacy systems
  - Data is generally of good quality at Kela with everything in a single data warehouse
  - However many systems in use were made before the time of advanced analytics
- Getting the results to real use
  - Deployment to scale
  - Training the staff to use the results

What's next

- We will organize Kelahack: Data Science 2020, the first Kela data hackathon this spring for students
- Seminar on Ethics of AI, organized by Kela in February

Questions, comments?

Aaro Viertiö
Product Owner – Data Science
Kela
aaro.viertio@kela.fi

Kela|Fpa