Data-driven excavation damage inspection

Agentschap Telecom
Ministerie van Economische Zaken

Summary

As part of the Ministry of Economic Affairs and Climate Policy, Agentschap Telecom (the Netherlands Radiocommunications Agency) is not only an implementing body but also the regulator for telecommunications legislation and regulations, and is likewise responsible for supervising the Act on the exchange of information on above-ground and underground networks (WIBON). The purpose of this Act is to prevent danger or economic damage caused by damage to underground cables and pipelines. Each year, Agentschap Telecom carries out hundreds of inspections.

To reduce the amount of fibre-related excavation damage, Agentschap Telecom launched a pilot in 2020 to make information from a number of relevant data sources (cartographic data, area polygons, Twitter information, and so on) available interactively to its inspectors via a Shiny app, supporting them in carrying out their inspections.

In addition, a machine learning model was developed that predicts the likelihood of excavation damage on the basis of historical data at the moment a KLIC notification (excavation notification) is submitted to the Land Registry (Kadaster). This prediction is also available in the Shiny app.

In view of the strong substantive results of the prediction model, Agentschap Telecom, Full Orbit, DICTU and Oracle joined forces in a follow-up step to examine whether it would be possible to make the model's current set-up more efficient — with a view to future operationalisation — by using OML4R, Oracle's in-database machine learning technology. The results were highly promising, both as regards reducing the time needed to train the machine learning algorithm and in eliminating manual steps when making new data available.

In future, consideration will be given to how the prediction model can be further geo-operationalised and embedded in working processes, so that it can genuinely be deployed in carrying out excavation inspections.

About Agentschap Telecom

Agentschap Telecom is part of the Ministry of Economic Affairs and Climate Policy. It is both an implementing body and the regulator for telecommunications legislation and regulations, including for the Caribbean Netherlands (Bonaire, Saba and Sint Eustatius).

The agency does this in other fields too, such as the Act on safe excavation (WIBON), weighing instruments, gold, silver and platinum, and space activities.

Excavation damage legislation

Agentschap Telecom supervises the Act on the exchange of information on above-ground and underground networks (WIBON). The purpose of WIBON is to prevent danger or economic damage caused by damage to underground cables and pipelines.

Preventing excavation damage is a task for clients, excavating parties (those under whose responsibility excavation work is carried out) and network operators, and is achieved by sharing information and making arrangements. Within the framework of this Act, Agentschap Telecom carries out hundreds of inspections each year on the more than 700,000 mechanical excavation works that take place.

Excavation damage

Despite the exchange of information intended by this Act, excavation damage causes considerable (social) disruption and entails high direct repair costs: over €38 million in 2020. Moreover, in some cases it can lead to dangerous situations. In recent years this has been true in particular of fibre installation work in which gas pipelines are struck.

1 Selection from the infographic "WIBON en schade door graven", Agentschap Telecom, 31-5-2021

Anyone carrying out mechanical excavation work is obliged, before the work begins, to submit a so-called KLIC notification (excavation notification) to the Land Registry, including details of the excavation polygon: a map showing the place where excavation is to be carried out. The Land Registry then obtains data on the position of the underground networks within the excavation polygon from the network operators concerned and sends this area information for the specified excavation area to the party submitting the notification. Among other things, the area information states the date on which the excavation work will start, the details of the excavating party and the excavation polygon within which excavation will take place.

> *"The high economic damage caused, combined with being able to draw on a large number of relevant data sources, made fibre-related excavation damage a relevant subject for us to start a pilot on. The aim of this was to use data analysis to reduce the number of excavation damage reports"*
>
> **Dr Ryanne van Dalen**
> Data Scientist
> Agentschap Telecom

Step 1: Pilot to reduce excavation damage

To reduce the amount of excavation damage, Agentschap Telecom started a pilot in 2020 to bring together a number of relevant information sources relating to fibre installation and present them in an interactive web application based on the R Shiny framework. Through this app, all relevant data can be accessed at any time by inspectors on a laptop, tablet or mobile phone for the purpose of carrying out inspections.

Data elements

The following data elements are combined in the Shiny app:

- Current map data via PDOK (Publieke Dienstverlening op de Kaart). Through PDOK the Dutch government makes open datasets with up-to-date geo-information available.

- Information on the current installation of fibre throughout the Netherlands, obtained by way of the area polygons from the Land Registry.
- Aerial photographs and soil maps for visualisation.
- The GPS location of the inspector. One of the inspectors' wishes was to be able to see all current excavation notifications in the vicinity at a glance while out in the field. It can happen that another KLIC notification becomes active nearby after an inspection has been carried out. Supervision of excavation activity can be maintained more efficiently in this way.
- Current Twitter messages about fibre. The independent site Glasvezelgids.nl reports on developments in the field of fibre internet for consumers and businesses in the Netherlands. This information is relevant to the inspectors, because reports from the public can also be grounds for closer inspection.

Prediction model

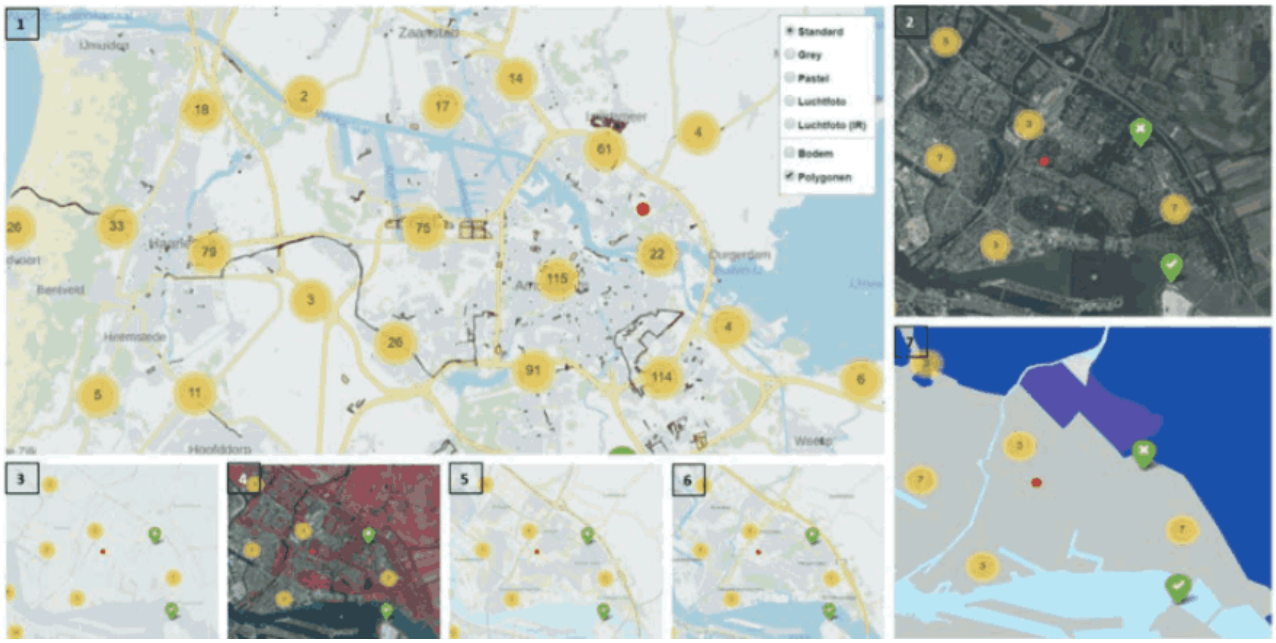
Alongside the development of the Shiny app described above, Agentschap Telecom also wanted to investigate whether a machine learning model could be developed that, on the basis of the information supplied when applying for a KLIC notification, could predict the likelihood of excavation damage.

In developing this machine learning model, five classification prediction techniques were applied to historical Land Registry data with different parameters. In the end the "random forest" algorithm was used, with the interpretability of the prediction model (explaining it to third parties) in combination with its predictive value proving decisive.

The most important predictors in the model are:

- Percentage of excavation damage per excavating party (number of excavation notifications with damage / total number of excavation notifications);
- Number of coordinates of the excavation polygon;
- Number of themes of the excavation notification (heat, water, data transport, etc.);
- Number and type of work in the excavation notification (pile driving, pressing/drilling, soil remediation, etc.).

In the end the accuracy of the model proved high: 82% of all predictions are correct. The PPV (the proportion of predicted excavation notifications with damage that are correct) is also fairly high: 62% of the correctly predicted damage is also actual damage. This is an important indicator for 'risk-driven' supervision. If inspectors are going to use the outcomes of prediction models to prioritise their work in future, you want them to visit excavation notifications where there is little chance of them coming for "nothing".



2 Example of a number of data elements displayed in the Shiny app

The outcome of the prediction model is a probability of excavation damage for every excavation notification, between 0 and 100%.

Step 2: Increasing the efficiency of the prediction model

Agentschap Telecom ran the pilot on a shared server with R and RStudio in order to model the data and drive the random forest algorithm. In view of the strong substantive results, it was decided to set up a follow-up step in the form of a POC, with the aim of examining whether the prediction model could be made more efficient with future operationalisation in mind. The points of attention in this POC were:

- reducing the time required to train the random forest algorithm. In the pilot's original set-up, based on open source technology, training the model already took 6 hours. With the expected growth in data, this would only take longer.
- examining whether a set-up would be possible in which new data could easily be made available without manual steps such as exporting and local storage.

In order to optimise Agentschap Telecom's prediction model further and operationalise it in future, Agentschap Telecom, Full Orbit, Oracle and the ICT Implementation Service (DICTU) joined forces to experiment, by way of a POC, with running in-database machine learning based on Oracle Machine Learning for R (OML4R).

What is OML4R?

Oracle Machine Learning for R can be seen as an extension of 'open source R', in which R is supplemented with Oracle R. Besides all the open source possibilities that R already offers, OML4R provides a range of packages for data manipulation, data visualisation, statistics and in-database machine learning with data from an Oracle database. The great advantage is that you can use the computing power of the database, and the database is of course an ideal place in which to supply new data that is then immediately available for a predictive model.

Predicting excavation damage with OML4R

To achieve all this, the parties mentioned above first examined whether it would be possible to make the excavation damage predictions using OML4R. Full Orbit, DICTU (responsible for Agentschap Telecom's database servers) and Oracle jointly supervised and carried out the installations and looked into the impact of installing and using OML4R within the existing Oracle Exadata database landscape.

The results were highly promising! The possibilities for in-database data manipulation and preparation using the Oracle R packages were an adequate alternative to the set-up in which the open source R package tidyverse was used. The follow-up step, running the in-database random forest algorithm, produced substantively comparable results to the initial set-up. In addition, the great advantage lay in the run time for training the random forest model: in the current set-up this takes 6 hours, whereas with a comparable dataset the run time was reduced to 2 hours using OML4R.

> *"Our aim in carrying out the POC was to see whether we could improve the performance of the prediction model, reduce the number of manual steps and consider whether the model could be incorporated into our BI landscape"*
>
> **Mustafa Bozkurt**
> Senior Data Consultant
> Agentschap Telecom

The future

On the basis of the good results of the prediction model, the follow-up step will consist of examining how the model can genuinely be geo-operationalised. The points of attention here are the technological integration of the model into the existing business intelligence landscape, as well as the

creation of sufficient support among the intended end users, so that they have the confidence to trust the model's outcomes and will use them when carrying out their inspections.

Find out more

Would you like to know how you too can make more impact with data science in your organisation? Then please do get in touch with us. We would be glad to tell you more.

Get in touch

Full Orbit
Jan van Krimpenweg 7-Q
2031 CE Haarlem

E info@full-orbit.com
T +31 (0)20 5300 390

Socials: