Transport for London Uses RapidMiner to Aid the Performance of the Road Network for Everyone

Executive Summary

- Transport for London (UK) is adopting an increasingly data-driven approach to road network management
- But historical approaches to data prep were manual and time-consuming
- RapidMiner’s automated workflow has streamlined the process - in one case converting a 42-step, full day manual process to a single click
- Now Transport for London can aggressively pursue new data sources, build more models and run them more often - improving city traffic flows
- RapidMiner will also help Transport for London bridge its skills gap and empower more employees to perform data science

About Transport for London

And the RapidMiner user group

Transport for London (TfL) is the integrated transport authority responsible for keeping London moving and the day-to-day operation of London's public transportation network and the management of London’s main roads. The transportation network includes London Underground, London Overground, Docklands Light Railway, TfL Rail, Trams, buses, taxis, cycling and pedestrian provisions, and river services. The underlying services are provided by a mixture of wholly owned subsidiary companies (principally London Underground), by private sector franchisees (the remaining rail services, trams and most buses) and by licensees (some buses, taxis and river services).
The RapidMiner users at TfL are part of Operational Analysis, a data science and analytics team within the Network Performance department. Network Performance is responsible for the safe and efficient operation of the road network, managing the traffic signals and ensuring safe, high-quality roadworks across the city. Their mission also includes achieving more progressive goals such as increasing the usage and efficiency of sustainable modes of transportation (bikes and buses), and limiting real-world disruption by modelling and visualising future changes.

**The Challenge**

TfL’s need: make traffic optimization a more data-driven process

Operational Analysis supports Network Performance through managing, developing and exploring a range of data sources to provide additional insights to the operational teams managing the traffic signals. Their work involves capturing, cleansing, combining and analysing large quantities of data from all transport modes. Their outputs include periodic reporting on the performance of the road network, the use of real-time data feeds to display live information about the network in a variety of dashboards, and analysis to answer questions posed by teams across TfL. In addition, they continuously seek and utilise new data sources, such as freight vehicle telematics data and data from cyclist activity tracking apps.

With better data and insights provided by Operational Analysis, Network Performance can make more impactful adjustments to traffic lights and other decisions to optimize traffic in the city.

**Before RapidMiner: many manual steps to prep data for optimization**

In the past, Operational Analysis relied primarily on Oracle SQL Developer to query data sources, and Microsoft Excel to manage, transform and analyse data. But more recently, as data sets have grown in size, Operational Analysis have been using R to handle more data and build predictive models, so as to provide additional insights to the rest of TfL.

For example, Operational Analysis developed an algorithm to analyse the patterns in sequential journey times along a given stretch of road, in order to estimate the amount of delay being generated by different causes. The bespoke algorithm itself was written in R, but the process of extracting, reformating and manipulating the data to be fed into it, and of aggregating and displaying its results, was a manual and time-consuming process using many different software packages. Running the whole 42-step process would take one of the team’s data scientists a full day of effort.
The Solution and Results

**RapidMiner's workflow turns 42 manual steps into one click**

With RapidMiner, Operational Analysis was able to build an automated data prep workflow using RapidMiner's intuitive, graphical user interface. For example, each of the 42 steps in the data prep process mentioned earlier was built into a workflow, so that when one step is complete the next one is triggered automatically. So, not only are Operational Analysis staff spared from performing each step themselves, they don't even need to kick off each successive step. It's a case of “set it and forget it” - until the data prep is complete.

**RapidMiner increases Tfl’s analytical bandwidth and capacity for traffic optimization**

RapidMiner has increased Operational Analysis’s capacity to support the wider business’s work, and more importantly, to improve the travel experience of London's residents and visitors. With RapidMiner replacing so many manual processes, the team is able to pursue new data sources - such as anonymized cyclist data from Strava, on-board diagnostics parameter ID (PID) data from vehicle manufacturers and others in the works. The team can also build more models and run existing models more often. These all result in more data and more insights which improve the optimization of London traffic flows.

**Future plans: RapidMiner will help Tfl bridge its data science skill gap**

More predictive models are in Tfl’s future. But like many organizations, Tfl has a skills gap - more demand and interest in data science than supply of trained professionals on staff. And making everyone learn to code in R is not a practical solution. RapidMiner will help. Tfl is starting to use RapidMiner not only for data prep, but for building models, too. RapidMiner’s easy-to-use interface, wizards and automated features make it the perfect solution to enable Tfl staff without data science backgrounds to perform true data science work. As this happens, the understanding of how the road network is performing will become even greater.