The Alan Turing Institute

Department for Transport
Natural Language Toolkit

TIN-DFT-001

About the Organisation

The Department for Transport (DfT) plans and invests in plan and invest in transport infrastructure to keep the UK on the move: via rail, buses, road, cycling, walking and more. We work with our agencies and partners to support the transport network that helps the UK's businesses and gets people and goods travelling around the country.

Role Description and Responsibilities

You'll join DfT's Data Science team: a friendly and capable team of data scientists with a range of backgrounds and career stages. Our projects cover a range of policy areas and internal organisational needs. We are ambitious about developing our skills, expanding our reach and increasing our impact.

We are looking for a Natural Language Processing (NLP) experienced intern to design and deliver an NLP toolkit for internal use. Work has already begun to develop and deploy some elements of the tool, but there is huge scope to increase the useability and content of the tool. We hope for example to include both sentiment and topic analysis, to create a tool that would drive efficiencies in policy making and the departments operations. This will include both a user interface as well as some level of automating outputs.

When complete the tool will be accessible to everyone within the organisation on a self-serve basis. Typical use cases of the tool are analysis of free text responses to public consultations and surveys.

Responsibilities include:
• Reflecting on the user research and progress so far
• Planning the next stage of development through to completion. Including:
   o Creating a user interface
   o automating the cleaning of data and,
   o creating useable outputs for policy professionals.
• User testing and incremental changes.
• Working alongside colleagues in digital services to deploy the tool on Google Cloud Platform (GCP).

There may be opportunities to work on other data science projects whilst with the team.

Expected Outcomes

During the internship you will deliver an NLP toolkit and deploy it on GCP for use by colleagues across the organisation. The toolkit must overcome the challenges identified in the user research. Following deployment, you will ensure the tool is used effectively to make the analysis stage of the consultation process more efficient.

Supervision and Mentorship

The project will be supervised by a Principal Data Scientist in the team and you will be supported by colleagues with the data science team of the Advanced Analytics Division.

Ideal Intern

To succeed in this role, the candidate should have:

- Practical experience in python
- Knowledge of NLP approaches and limitations
- Experience of deploying models on GCP and willingness to explore new tools that may be suitable. The tool currently uses VertexAI pipelines, AutoML, cloud functions, and cloud storage.
- Practical experience in using specialised machine learning libraries e.g. sci-kit learn
- Strong verbal/written communication & data presentation skills are recommended.

Internship Logistics

Start date: September 2022 (negotiable)
Duration: 3-6 months (the length of the internship is negotiable)
Location: Hybrid working based in London office (>40%) and working from home
Full time or part time
Renumeration: £30,000 p/a pro rata