

# DSC-172 Optimus 2 - classifier

2019-08-01

Summary - Following on from optimus, several Government departments have labelled ferry data. This project aims to take those labelled datasets and build a classifier.

## Team members

- Steven
- Evie

## Impact

The project is high priority to several government departments such as BEIS, DIT, DFT and DEFRA. It is also included in the DIT work package.

**Which of the Campus strategic objectives would this project deliver, and how?** \* Policy Impact \* Technical Impact

**What is the policy impact? Why is this project important?** \* Understand trade implications

- How will operational efficiency or value for money be improved? What will be the impact of this?

Make unusable datasets usable by allowing aggregation of free text.

- What is the impact of not doing this on our reputation?

There is a big demand for this across several departments. If we didnt do this we would significantly damage our reputation, especially after having delivered the initial tool which is held in high regard among the customers.

- How will learning be shared beyond the initial project? (e.g. through re-use, understanding data, best practice etc.)

**What is the technical impact?**

\*Experimenting with classifiers on truly free text inputs. Most nlp methods require data cleaning and structuring before training. This model is exploring building classifiers directly on messy natural free text data.

- What is the data science you are expecting to use, and why is this interesting? Character embedding. Neural Nets and maybe other classifiers like Random Forests etc. . .
- Will the data use new or novel data sources, or use existing datasets in novel ways?

New data source with government departments wanting to give us more data.

**Why should the Campus do it?** High political and reputation benefit

### **Code and outputs**

- Processed data sets for government trade analysts
- We can also give this model to other departments to apply to future data provisions

### **Related and existing work**

- Optimus

### **Delivery**

ASAP

### **Updates**

- No updates yet.