

Novel approaches to the Living Costs and Food Survey

2018-07-17

This project aims to explore the application of computer vision and Natural Language Processing (NLP) techniques to the Office for National Statistics (ONS) Living Costs and Food Survey (LCF). Specifically, we will produce a set of tools for automatically extracting textual data from scanned shopping receipts (optical character recognition, OCR) and then convert this unstructured text data into tabular form using various NLP techniques.

Team members

- Lan Benedikt
- Ian Grimstead
- Alex Noyvirt
- Jeremy Rowe
- David Pugh
- Sharon Hook

The need

Explore the use of receipt scanning data and barcode scanning data to replace manual data entry of the LCF diaries. The purpose is to reduce respondent burden and make efficiency savings, we will benchmark the automated process against the current manual process. Success measures are speed and data quality.

Impact

There is a clear goal to improve a production process. There is also a possibility for knowledge sharing with other National Statistic Institutes (NSIs) and the opportunity to reuse some of other Data Science Campus projects, such as Optimus.

Data science

- Image processing
- OCR
- Supervised, unsupervised machine learning (ML)
- NLP
- Data linking

Stakeholders

- LCF team in Social Survey Division
- ONS Prices Division
- Statistics Netherlands
- Statistics Austria
- Statistics Finland
- Statistics Slovenia

Further information

Please contact datasciencecampus@ons.gov.uk for more information.

Updates

- No updates yet.