DSC-24 Classification of financial services

2018-04-12

This project explores whether it is possible to classify financial corporations to their detailed [Standard Industry Classification 2007](https://www.ons.gov.uk/methodology/classificationsandstandards/ukstandardindustrialclassificationofeconomicactivities/uksic2007) (SIC2007) using data on their financial assets and liabilities, and other firm-level information. The project makes use of a number of unique features derived from both structured and unstructured text.

## Team members

* Alex Noyvirt

## The need

The ability to classify the financial companies into sub-sectors by using firm-level information on their financial assets and liabilities, and other business information such as turnover and employment, has been identified as a high priority for the Flow of Funds collaboration project between the Office for National Statistics (ONS) and the Bank of England (BoE).

## Impact

The outcomes:

* the recommendations made by the report can be used to improve the financial services survey by designing additional questions with more discriminative power for classification
* companies highlighted by the algorithm as potentially misclassified by the SIC 2007 code recorded in the Inter-Departmental Business Register (IDBR) can be included in the process for manual revaluation of their activities
* methodology can be reapplied easily for classification of companies from other administrative datasets

## Data science

Machine learning, that is, Random Forest and XGBoost in a distributed environment (Cloudera and Apache Spark) have been applied together with an extensive set of feature selection methods. The work has been implemented using Scala.

## Stakeholders

ONS flow of funds - economic statistics

## Delivery

* [x] February 2018: feature selection methods developed
* [x] May 2018: processing of 6 million machine models
* [x] June 2018: evaluation of the results
* [x] June 2018: project report work
* [x] August 2018: report finished
* [x] September 2018: show & tell to stakeholders (FCA)
* [x] Future: project completed

## Further information

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

## Updates

**2018-11-29T15:44:43Z**

[Alex Noyvirt](https://datasciencecampus.ons.gov.uk/author/an/) has published a [blog post](https://datasciencecampus.ons.gov.uk/finbins-granular-classification-of-the-uks-financial-sector/), [technical report](https://datasciencecampus.ons.gov.uk/wp-content/uploads/sites/10/2018/11/Technical-report-Machine-learning-for-classification-of-financial-services-companies-in-the-UKs-financial-s.pdf) and associated (Scala) [code](https://github.com/datasciencecampus/finbins) for this project.