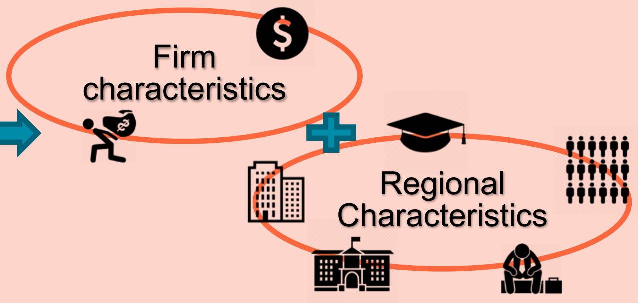
Prediction of Company Success/Failure Using Supervised Machine Learning Algorithms

Project Goal

Create a ML algorithm that predicts success or failure of new tech businesses in the UK from



"Machine learning is based on algorithms that can learn from data without relying on rules-based programming."- McKinsey & Co.

Project Execution

Data Gathering

- **Business dataset**: Open source dataset containing all registered businesses in the UK.
- **Census data**: open source census data at Ward level
- **Postcodes data:** Dataset containing postcodes and ward codes to make it possible for data sets to link

Data Processing

Data was filtered to:

- Tech companies
- Opened within the last 15 years

Data was processed so that each variable is assigned to its right class.

Data was sub-setted to include only variables of interest.

Data Linkage

Business Data + Postcode data

By Postcode

Linked data + Census data

By wardcode

Data was manipulated to include % of unemployed population, % of population with higher education degrees

Data Analysis

- Data was split into Training dataset and Testing dataset

Machine learning algorithm used and compared:

- Logistic regression classifier
- Random forest classifier

The algorithms were trained on training data and accuracy was tested on testing data set

Key Findings

Logistic regression found that significant variables in predicting company closure were:

- Number of mortgages outstanding
- Percentage of economically active population
- Percentage of unemployed population
- Percentage of population with a higher education degree
- Number of companies in the city
- Number of universities in the city

Unbalanced Nature of data required adjusting the logistic regression classifier and the use of random forest classifier

- Changed classification threshold from 0.5 to 0.10, 0.15 and 0.20
- Downsized the dataset to make it balanced

When comparing the accuracy of the models, it was found that random forest classifier and logistic regression (0.10 threshold) where the best in predicting company closure from the significant variables

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

The University Of Sheffield.

Gioia Iacopini
giacopini1@Sheffield.ac.uk