Classifying Illness Complexity

John Bell
MSc in Big Data

Overview

The purpose of the project is to build a machine learning classifier to categorise health-care users in Scotland based on the over-all complexity of their health conditions.

The output of the classifier will be incorporated into the High Health Gain (HHG) model (which predicts the top 5% of patients by resource utilisation) as a risk factor.

The host organisation for the project is NHS National Services Scotland, Information Services Division (ISD).

Background

Comorbidity (i.e. two or more conditions being present in an individual) is a major issue in public health, and is strongly associated with poorer health outcomes and poorer quality of life. It is a key factor in determining the complexity of an individual's health conditions. A means of classifying the complexity of patients health (using comorbidity and other factors) does not currently exist for the Scottish population.

Approach (using CRISP-DM)

Business Understanding: undertake a literature review of widely-used comorbidity measures (i.e. statistical tools which are used to predict mortality, clinical outcomes or health-care costs). Work with clinical experts to understand the health-care context and inform a suitable approach.
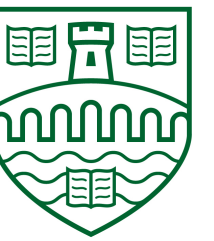
Data Understanding: undertake EDA. Work with information analysts to understand the data.

Data Preparation: aggregate patient-episode level data to patient level across acute hospital admissions, outpatient attendences, A&E, geriatric and pyschiatric datasets. Integrate Pharmaceutical datasets. Prepare time-based queries for 'snapshotting' a patient's condition at a given point in time.

Modelling: build logistic regression model as baseline, then compare random forest and neural network.

Evaluation: evaluate the classification models against unseen data. Evaluate chosen classification model(s) against HHG. Communicate results and ideas for future development to internal stakeholders.