Data Analytics Coursework 2 - [SET09120](https://moodle.napier.ac.uk/course/view.php?id=35533)

**Davide Pollicino**

40401270@live.napier.ac.uk

Edinburgh Napier University, Edinburgh (UK) - [SET09120](https://moodle.napier.ac.uk/course/view.php?id=35533)

Abstract:

Keywords: Open Refine, weka, data preparation, data cleaning, python, regression, classification, j48.

1. Tools Used and Approach to the coursework

In this coursework, I have been using Google OpenRefine and Weka. Google OpenRefine,  is a standalone open source desktop application for data cleanup and transformation to other formats. t is similar to spreadsheet applications; however, it behaves more like a database.

Weka is Environment for Knowledge Analysis, developed at the University of Waikato, New Zealand, is free software licensed under the GNU General Public License, and the companion software to the book "Data Mining: Practical Machine Learning Tools and Techniques"

2. Data Preparation

2.1. Data Cleaning Describe the way the data is cleaned by the use of OpenRefine, including errors identified and corrected.

2.2. Data conversion Describe how the cleaned data is converted to data sets, which can be analysed by algorithms in Weka.

3. Data Analytics

3.1. Classification

3.2. Regression

3.3. Association

3.4. Clustering

Note: • Three of the above 4 sections will be fine. For each section, apply at least one of the algorithms in that category. Settings, datasets used, attributes involved for each run should be described. Patterns/rules need to be retrieved, analysed and interpreted. About 6 patterns/rules for each algorithm used are expected.

4. Summary/Conclusion References are expected