

**ECOS Project: Using Data Analysis to Improve
Marketing Response at D.C. Thomson**

Tom Davenport

September 2016

**Dissertation submitted in partial fulfilment for the degree of
Master of Science in Big Data**

**Computing Science and Mathematics
University of Stirling**

Abstract

Data analysis tools and techniques provide a substantial opportunity for business. Companies have become able to store more information not only about user's personal information, but also around their interaction with the company, through their response to emails and to how they behave on their site. This has created a vast pool of data that when analysed appropriately can provide useful evidence for marketing decision making. The data however needs to be put through a series of processes to ensure that it is of a sufficient quality that analysis is grounded in fact and will be of business value.

This project was undertaken for the DC Thomson company to examine the response to emails sent during the Christmas season from October to December of 2015. It created a series of data models that would highlight what important factors were involved in making an email successful, and whether there were patterns amongst groups of users that purchased from emails. The objectives of the project were to create one model that could reflect the structural factors that made emails successful, and another model that would show what factors had an influence in leading a customer to purchase. It involved a multi-level analysis of different sources of data, creating and assessing the quality of datasets that could be used to generate data models.

The approach to modelling involved examining current methods of email data mining, and following the CRISP DM standard for approaching data mining projects. The development of the project was undertaken within an Agile environment, using daily stand up meetings and weekly Sprint meetings. Different methods for creating the analysis were considered, with the majority of data cleaning being undertaken in R, data scraping and merging using Pandas in Python. The final models were developed using R. These models generated were decision trees, a white box model that can effectively show the relationships within the data that lead to a particular outcome.

The results of the project were presented to senior management. This included suggestions of how they should test campaign emails in future, by what structural influences the model had suggested generated a good response. An audit of missing data from the user campaign was also presented, showing the importance of each variable and its proportion of missingness from the set.

The process of building these models was documented effectively which will allow them to be recreated and extended with further detail to study other campaigns. The suggestions from the models will have an impact on the organisation's decision making process in the future.