

# TEXT ANALYTICS USING RAPID MINER

By

Aaron Ward

Supervisor(s):

SUBMITTED IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF  
B.SC IN COMPUTING AND INFORMATION TECHNOLOGY  
AT  
INSTITUTE OF TECHNOLOGY BLANCHARDSTOWN  
DUBLIN, IRELAND  
2017

## **Declaration**

I/We declare that this material, which I/We now submit for assessment, is entirely my/our own work and has not been taken from the work of others, except where otherwise stated. I/We have identified and included the source of all facts, ideas, opinions, and viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the source cited are identified in the assignment references.

I/We understand that plagiarism, collusion, and copying are grave and serious offences and accept the penalties that would be imposed should I/we engage in plagiarism, collusion or copying. I acknowledge that copying someone elses assignment, or part of it, is wrong, and that submitting identical work to others constitutes a form of plagiarism. I/We have read and understood the colleges plagiarism policy 3AS08.

This material, or any part of it, has not been previously submitted for assessment for an academic purpose at this or any other academic institution. I have not allowed anyone to copy my work with the intention of passing it off as their own work.

Dated: 2017

Author:

---

Aaron Ward

# **Text Analytics Using Rapid Miner**

## **Business Understanding**

This sections shall describe the business objectives, the mining objectives, a analysis and plan for the proposed project.

### **Business Objective**

The objective of this project is to mine non-structured data using Rapid Miner.

### **Data Mining objective**

- The main objective is to mine articles about Machine Learning, Deep Learning and Robotics.
- Webcrawling will be implemented to find related text on the topics.
- Multiple clusstering and classification models will be used to identify the texts.

**Project analysis**

**Project plan**

**Data Understanding**

**Data Preparation**

**Modeling**

**Modeling techniques**

**Clustering**

**Classification**

**Test Design**

**Build and Assess the Model**

**Evaluation**

**Project review**

**Project deployment**

# **Appendices**

## **Appendix A**

### **Rapid Miner Processes in XML**