**Fake News Detection: Product Backlog**

|  |  |
| --- | --- |
| Report Name | Product Backlog |
| Author (User Id) | William Minton (wim12) |
| Supervisor | Richard Jensen (rkj) |
|  |  |
| Module | CS39440 |
| Degree Scheme | G400 (Computer Science) |
|  |  |
| Date | February 16, 2019 |
| Revision | 0.1 |
| Status | Draft |

# User Stories

1. Research and Investigation into previous attempts at similar work – Books, Scientific papers and API’s for third-party libraries
2. The system should use a version control system such as github, to monitor progress and serve as a way of backing up data
3. The system should be functionally comparable to other fake news detection systems, and be compared to existing methods of fake news detection such as knowledge engineering, stance detection or metadata comparison
4. A dataset should be researched and decided on – Possibly multiple
5. The system should work with a wide-ranging, functional dataset, although should be constrained to a 2-class classification problem
6. The Scikit-Learn library will be used and researched to assist with this
7. The Bag-of words approach of feature extraction will be used
8. The TF-IDF approach of feature extraction will also be used
9. The titles and text of the dataset used will be cleaned using regular expressions

Four machine learning algorithm models will be applied to both methods of feature extraction:

1. Multinomial Naïve-Bayes
2. Linear SVM
3. Ridge Classifier
4. Decision Tree

If there’s time, other methods of machine learning will be applied

1. Convolutional Neural Networks using Keras
2. K-Nearest Neighbours
3. ???

Various different evaluation metrics will be used for evaluation and comparison:

1. Accuracy
2. Precision
3. Recall
4. F1 Score

Various different methods will be used to visualise and compare this data:

1. Tables
2. Confusion Matrices
3. ROC curves

Written Deliverables:

1. Project Outline should be worked on
2. Ethics form should be submitted
3. The mid-project demonstration should be worked on and practiced
4. Presentation for the mid-project demonstration using powerpoint/Openoffice impress
5. Script written for the mid-project demonstration
6. The final demonstration should be worked on and practiced
7. Presentation for the final demonstration using powerpoint/Openoffice impress
8. Script written for the final demonstration
9. Documentation written on the installation and use of the code provided
10. Annotated Bibliography should be worked on over the length of the project
11. Final Project Report should be worked on over the length of the project

# Sprint Backlog (Basic)

## Sprint 1 – 1/2/2020 – 10/2/2020

Item 1 – In depth

Item 6

Item 20

Item 24

Item 33

## Sprint 2 – 12/2/2020 – 19/2/2020

Item 2

Item 4

Item 7

Item 8

Item 9

Item 10

Item 22

## Sprint 3 – 22/2/2020 – 2/3/2020

Item 11

Item 17

Item 18

Item 19

Item 20

Item 25

Item 26

Item 27

Item 28