Topic Classification in Social Media

PSG 16 – Dr Zhiwei Lin – BEng Software Engineering

Thomas Franklin B00607399

April 2018

# Abstract

In recent times there has a been a rise in what is deemed as “fake news” when it comes to social media sites such as Twitter. Fake news can be described as the spreading of misinformation to influence the readers thoughts and is becoming a much larger issue in recent times with the rise of social media and particular “influencers”.

As social media tends to be an area where users can express their own belief, large influencers can use it as a playground to misinform their followers and hence start the process of spreading fake news – since content on social media can reach hundreds of thousands of users in minutes, which means current methods of combating fake news is often to slow to stop the spread before it starts, it was time to look in to solutions which can aid in detecting fake news a lot faster and prevent the spread before it is too late.

The aim of this project was to use machine learning to classify all public Tweets from Twitter as a rumour or non-rumour where the results could then be retrieved through a rich internet application for visualisation. From the research conducted there were very few existing projects that have tackled rumour detection in social media, and from the information gathered this will be the first project which uses natural language processing and machine learning for the early detection of a rumours in social media.

The final product produced is a service which continually streams Tweets from Twitter based on keywords, puts them through a processing process which classifies them as either as rumour or non-rumour which is then stored in a SQL database for later retrieval on an interactive web service which will show a breakdown of the results in graphs, charts etc.

In conclusion, the use of natural language processing and machine learning is only the start of tackling fake news, on its own it reduces the amount of results which will still need accessed, as the current methods where individuals who cross-verified results having to go through everything – by reducing the results earlier in the process it can speed up the cross-checking steps that will need to be done.

# Acknowledgements

This project would not have been possible without the support of my mentor Dr. Zhiwei Lin for his guidance and support throughout. Another thanks goes towards the Ulster University teaching staff of whom I had the pleasure to learn from over the last 4 years, and finally I would like to thank those I had the pleasure of working with while on placement – most importantly Aaron Long who was a mentor to me when starting out in placement and with his continued support I have become a better developer.

**Table of Contents**

[Abstract 1](#_Toc509320265)

[Acknowledgements 1](#_Toc509320266)

[1. Introduction 3](#_Toc509320267)

[1.1 Problem Elucidation and Statement 3](#_Toc509320268)

[1.2 Project Aim 3](#_Toc509320269)

[1.3 Project Objectives 3](#_Toc509320270)

[1.4 Selected Software Lifecycle Methodology 3](#_Toc509320271)

# Introduction

This report will outline the solution developed to tackle the problem identified in the report, along with describing the development process of the proposed solution along with the management and elicitation of the product and user requirements which will be documented in the report, the report will also highlight the design and testing which had been completed to ensure that the project is a success.

## Problem Elucidation and Statement

## Project Aim

## Project Objectives

## Selected Software Lifecycle Methodology