

# PROCESSING WHAT MAKES A GOOD REDDIT COMMENT

ANDREW McLEMAN, JOE McDONALD, JORDAN YOUNGMAN,  
MURRAY LYNE, SCOTT THOMSON



A REPORT SUBMITTED AS PART OF THE REQUIREMENTS FOR THE DEGREE  
OF MSc IN CYBER SECURITY  
AT THE SCHOOL OF COMPUTING  
ROBERT GORDON UNIVERSITY  
ABERDEEN, SCOTLAND

April 2019

Supervisor Eyad Elyan

# Abstract

The Abstract of the report should be written here, it should provide a short summary of the work encompassing no more than 300 words.

# Declaration

I confirm that the work contained in this MSc project report has been composed solely by myself and has not been accepted in any previous application for a degree. All sources of information have been specifically acknowledged and all verbatim extracts are distinguished by quotation marks.

Signed .....

Date .....

Andrew McLeman, Joe McDonald, Jordan Youngman, Murray Lyne, Scott Thomson

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Declaration</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 About this Thesis . . . . .	1
1.3 Conclusion . . . . .	1
<b>2 Background Research</b>	<b>2</b>
2.1 Political Motivations behind Subreddits . . . . .	2
2.2 Psychological Traits in Users . . . . .	2
2.3 Significant Events . . . . .	4
2.4 Machine Learning . . . . .	4
2.5 Conclusions . . . . .	5
<b>3 Method</b>	<b>6</b>
3.1 Conclusions . . . . .	6
<b>4 Results</b>	<b>7</b>
4.1 Conclusions . . . . .	7
<b>5 Discussion</b>	<b>8</b>
5.1 Conclusions . . . . .	8
<b>6 Conclusion</b>	<b>9</b>
6.1 Conclusions . . . . .	9
6.2 Future Work . . . . .	9

# List of Tables

# List of Figures

# Chapter 1

## Introduction

A short paragraph introducing the topic the chapter examines.

### 1.1 Background

A number of pages about the background of the project.

### 1.2 About this Thesis

This is the thesis of *Insert Full Name Here*, submitted as part of the requirements for the degree of MSc Computing: Software Technology at the School of Computing, Robert Gordon University, Scotland.

A number of paragraphs detailing the main expectations of this body of work.

### 1.3 Conclusion

A short conclusion summarising the chapter.

## Chapter 2

# Background Research

This chapter provides some background research on the project and examines some previous work.

### 2.1 Political Motivations behind Subreddits

Reddit is home to several political subreddits, many of which have high subscriber counts and activity per day. As stated in the paper titled Automating power political actors have used bots in social platforms to influence public opinion. They are used to not only raise awareness in certain political campaigns but also to follow politicians on platforms such as twitter and Facebook to give the illusion of popularity. They can even do an attack on news outlets where they flood their wall with misinformation as to avoid and sidetrack public attention. [10] the above research strongly shows why in our paper it is imperative that we distinguish AI from human users to get accurate and truthful data to work with. Also stated by the paper titled Melanization of Politics it describes and researches in the rise of political discussion due to the rise of media in the late nineteenth century early two-thousands it discusses how politics is now spoken on such a large scale. [11]

### 2.2 Psychological Traits in Users

Psychological traits play an important part in social media posts.

This paper[1] suggested that Users with Stronger personality traits tend to have stronger engagements in online discussions and therefore have increased interactions on Reddit. These users tend to use social media more because they have a more extroverted personality, and are much more open on social media platforms. This also means



due to their extrovert traits, that these users tend to have better leadership qualities. The results, however, had issues with statistical significance.

Users tend to display “Cyborg-like” behaviour when trying to attract attention to their posts. This doesn’t always work and a significant number of posts studied [7] failed to garner the attention they wanted.

**User Behaviour Patterns** It was found that women tend to post far less than men, and age determines how common a user will post on the platform. Users who had higher news engagement tend to comment and vote more on news related posts. As well as that, Older users were much more likely to post on Reddit. Voting patterns have no significant predictors, with the only exception being that news engagement can predict voting. It was suggested that the reason women post less is due to the fear of suffering from online harassment, as Reddit has a history of harbouring subreddits that have “existed to only annoy other redditors”. Egregious examples include /r/fatpeoplehate and /r/rapingwomen. Of which, were banned by Reddit in 2015. [4]

When it comes to positive events occurring on Social Media, it was found that there is an increase in negative emotions in social media on Twitter. It was found that on Reddit, highly upvoted stories tend to gather a proportionate amount of downvotes, therefore indicating that the same effect can happen on Reddit as well.[3]

Most posts on Reddit “die” after one day in terms of activity, which is a pattern that has been observed on many other social media platforms. For posts that only had 1 posted comment, it was found that 72.78 Per cent of these post did not last more 600 seconds in terms of new activity [7]. Users who comment frequently on other user posts tend to have more highly scored posts and therefore if you want to have good interactions with other posts, you need to be reciprocative. 67 Per cent of authors had more effective comments than posts compared to 22 Per cent who has less effective comments than posts.

Most users are only active on a handful of subreddits. It was found that during a 1-year data collection on 309 Reddit users, only 109 unique users(104 subscribed and 44 unsubscribed) had subscription events fired (This is when a user subscribes to a subreddit). Users tend to have varying attention spans as 73 Per cent of posts are rated without viewing the content of the post first. It was found that most users were “headline browsers” who only look at the headline of a post before voting on it. It was also found that most users also only look at the top headline posts on the front page, which is known as position bias. This means that users tend to gravitate towards posts that are on top of the front page of a subreddit, compared to posts further down the page. Users often vote on posts before viewing the comments on the post. Over 50 Per

cent of users vote before actually checking the comments replying to the post. [8]

It has been noted that the variety of subreddits browsed by users is lacking. A severe lack of browsing and voting variety was noted which indicates that users tend to create “echo chambers”, in which they only view content that they agree with. [8] This is a very commonly observed phenomenon across many social media platforms, and Reddit is no exception. [8]

Moderation doesn’t always affect user behaviour on Reddit. Subreddits themed around real-world discussion tend to value analytical and objective based comments compared to less serious subreddits. Even when filters are enabled preventing certain types of content being submitted, this makes little difference in the behaviour patterns of users.

## 2.3 Significant Events

Governments use bot farms on social media to manipulate public perception on social media, and Reddit is no exception. Many Governments, militaries and significant state actors have been outed as having used social-bots to appear either more relevant and influential or more controversially to directly manipulate legitimate users opinions on political events and military regimes. [5]

## 2.4 Machine Learning

Reddit is host to multiple AI experiments as well as many bots to reply to incorrect post syntax. Reddit is commonly used due to having various different subreddits from politics down to humour subreddits. This makes for an excellent place for AI to learn through varying teaching methods such as reinforcement learning to learn from peers in the subreddit or to strive for higher upvotes than previous posts. [2] The popularity of Reddit posts can be hard to classify without analysing the content of the post first. It can also be hard to classify posts if the subreddit in question covers a broad subject. Post popularity is determined on the content of the post, when the post was posted and the particular subreddit it was posted on, which, can be regarded as the context for discussion. An example of this would be the subreddit /r/pics, which had a poor performance in error rate due to the broadness of the subreddit’s context. Whereas other subreddits follow more specialized trends which are easier to classify. This paper ran classification algorithms and found that it ran better on simpler datasets. It also found that features like a Reddit post’s title lacked “indicative” power on whenever it made it a popular post or not. It suggested that more features from the Reddit API would improve prediction ability. [6]

With that said, this paper [9] suggests that it is possible to predict the score of Reddit comments, even if the communities are loosely defined (in terms of context) and disorganized. It found of note that user flairs (which are often given by moderators of a subreddit) can be an excellent predictor of highly popular comments. The model could also retrieve, with good precision, the highest rated comments within a particular post. It found that lowly rated comments tend to contain outdated jokes or information, which by that point had been cycled by the community to the point of exhaustion. Machine learning on Reddit needs to take into account community-related factors into what makes a comment popular. Posts on expertise related subreddit communities tend to use more technical and analytical terms. Likewise, comments on news related subreddits tend to be less emotional and value analytical insight over emotional responses. Some subreddits can vary in language use, the /r/worldnews subreddit is more diverse lexically, whereas /r/worldpolitics tends to use more “netspeak” compared to other subreddits. Emotional comment preference can vary between subreddits, with some subreddits heavily disliking having emotion in their comments. Time can play an important factor in determining how popular a comment can get. Comments on communities such as /r/news tend to be much more popular compared if posted later in time compared to comments made earlier, whereas other communities show the opposite happening. Moderation doesn’t always affect user behaviour on Reddit. Subreddits themed around real-world discussion tend to value analytical and objective based comments compared to less serious subreddits. Even when filters are enabled preventing certain types of content being submitted, this makes little difference in the behaviour patterns of users. This means that moderation doesn’t necessarily need to be taken into consideration when ranking comments, as the user behaviour is always affected by a significant amount.

## 2.5 Conclusions

The main conclusions for this chapter.

## Chapter 3

# Method

### 3.1 Conclusions

## Chapter 4

# Results

### 4.1 Conclusions

## Chapter 5

# Discussion

### 5.1 Conclusions

## Chapter 6

# Conclusion

This chapter summarises the main outcomes and conclusions resulting from this body of work.

### 6.1 Conclusions

The main conclusions that may be drawn from the body of work.

### 6.2 Future Work

Further development that could be carried out in the future.