Social Media Bias Analyzer

Project Specification

Joshua Fitzmaurice
Department of Computer Science
University of Warwick

0.1 Glossary

Bias - within this paper, bias refers to the over-representation of specified labelled topics. e.g. feed consisting of a lot of sport posts, or news posts.

1 Introduction and Motivation

Bias in social media can easily be seen on anyone's social media. In fact, I can show this with ease by just taking a look at my Instagram's "Explore page".

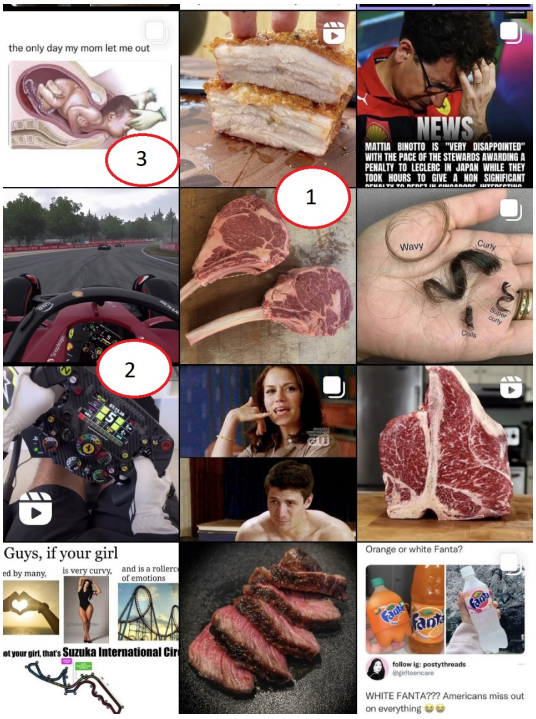

Figure 1: My Instagram for you page

Here we can notice a few common themes/biases: 1. Food, 2. Formula 1, 3. Memes. We want to be able to identify these biases for users so they can get an overview of the type of content they are receiving from social media.
With social media recommender systems programmed to entice users with content they will enjoy (Shin (2020)), it is common for similar groups of posts to be observed by a user if they have recently liked, commented, or viewed similar posts (Instagram).

2 Problem Statement

As discussed in Section 1, it is often the case that users are shown similar posts and not get a strong representation of posts from all aspects of social media. This in itself is not a major problem as users obviously want to see that content, hence why they like/view it. But it would be nice if users were able to see an analysis of the bias they observe in their social media feed (I know I want to see this information).

This project will involve creating a set of labels as well as training/identifying identifiers for said label. We can then scan through a users social media looking for the identifiers to analyse the type of posts prevalent. We can then further develop on this information by performing further data analysis (how, will be determined when patterns emerge in testing).
I have kept this description relatively concise and will develop further on it in my progress report/final report.

3 Objectives

Due to the late changes of project, the objectives/requirements are not fully complete. The goal of these objectives is to give a general understanding to any stakeholders in this project and not to enforce a rigid set of requirements for the implementation of this project.

3.1 Development of my own framework

This is the first, and primary, section of my project. It will involve designing and implementing a framework to detect and display over-representation of topics.

1. Be able to retrieve twitter homepage data from twitter API.
2. Generate a base list of topics we want to be able to detect.
3. Design a method of detecting different topics (here are some possible ideas)
    - Generate a list of keywords as "features"
    - Using the features we can train a ML model to predict the topic given a set of features.
4. Framework should be able to handles a set of posts (10-15) and for each post assign a topic it is representing
5. We can then use the results to perform further analysis.
    - This analysis will involve creating a set of rules for different social media accounts.

- Run the rules on different social media accounts and see how the bias changes over time.

6. EXTENSION - take into account images when analysing posts
    - First, use text detection to find text in the image, and add it to the post.
    - will require a method of object/item detection and labelling

3.2 Comparison of my framework to others

Once creating my framework, I need to analyse and compare it against others.

1. generate sample home twitter feeds.
    1.1. need to generate varying levels of bias within this dataset.
    1.2. generate erroneous test cases.
        - No matching keywords.
        - No posts given.
        - Post containing no text.
2. Determine accuracy of framework using sample twitter feeds.
3. Compare accuracy and other metrics with the given papers in Section 6.
4. Conclude the pros and cons of the different frameworks.

3.3 Data Analysis

This section describes what analysis could be done with the data as well as considerations needed to follow data privacy and data protection laws.

1. Determine a system for rule generation.
2. Setup social media accounts to test the rules on.
3. Analyse the data to determine the bias of the social media accounts.
4. Run the rules every hour on the top 100 posts of each account, analyising the new topical bias.

3.4 Chrome Extension development

After completing the analysis, I could, as an extension, create a chrome extension to display topical bias when on social media.
Functional Objectives:

- MUST
    1. Chrome extension must be visible when on Twitter
    2. Chrome extension must send post information as a request to API
        2.1. Scrape the first x posts from the homepage
        2.2. Using a GET/POST request, retrieve political alignment information from API
    3. API must be able to handle GET/POST requests giving post information
        3.1. receive a list of posts via a GET/POST request
        3.2. feed this list into the bias analysis framework
        3.3. return the corresponding results back to the chrome extension
    4. API must calculate the biases as per Section 3.1
    5. Chrome extension must display the bias information and further analytics via either numerical methods or visual methods.
- SHOULD
    1. Chrome extension/API should be able to handle search results as well as home page.
- COULD
    1. Chrome extension should be able to handle multiple social media sites

Non-Functional Objectives

- Chrome extension should update when social media site opened within 2 seconds
- Chrome extension should always appear to be updating/working
- Chrome extension should display when errors occur.
- Information displayed should be easily understood by the general public of the UK.

4 Testing

Different forms of testing will be used throughout the development of this project.
Black-/White-box unit tests will be created while designing/implementing the project. I plan on using test-driven development, so these tests will be necessary.
As well as this it will be important to test the framework as described in Section 3 to ensure we get useful information from the framework.
I will also include Integration/System testing for the software engineering part of the project (chrome extension).

5 Methods

5.1 Research

I will keep a written log of papers I read/use for this project as well as key areas of information found within the papers.

5.2 Technical Implementation

Python is the choice of language for the backend API as there are readily available libraries that provide the ability to create APIs, as well as access available social media APIs.
JavaScript is currently planned to be used for the Chrome extension.

The software methodology chosen for this project is a waterfall approach. I will not be sticking to a strict waterfall approach, however, as this could cause major disruptions in my time management if any changes to the requirements is needed. Any changes made to the specification during the development of the project must be added on in an agile-like manner to avoid missing deadlines.

6 Papers

In this section I will give a brief analysis of papers that attempt to achieve a similar goal to this project, and what useful information I have come across while reading these papers.

6.1 Pythia - Litou and Kalogeraki (2017)

Pythia is an automated system for short text classification. It makes use of Wikipedia structure and articles to identify topics of posts. Essentially, "Wikipedia contains articles organized in

various taxonomies, called categories". Pythia then goes on to use this information as their training data as well as handling sparseness in posts on social media.

6.2 Topic tracking of student-generated posts - Peng et al. (2020)

This paper proposes a solution for determining valuable information/topics discussed in student forums on online courses. It uses a model called "Time Information-Emotion Behaviour Model" or otherwise called "TI-EBTM" to detect key topics discussions , keeping in mind the progress of time throughout the forum.
Although this paper specializes in academic online forums, the approaches made could be relevant and useful for this project.

6.3 Topic classification of blogs - Husby and Barbosa (2012)

This paper uses Distant Supervision - 'an extension of the paradigm used by (Snow et al. (2004)) for exploiting WordNet to extract hypernym (is-a) relations between entitities' - to get training data via Wikipedia articles. Then trains their own designed model on this data to be able to classify topics via a multi-class recognition model (69% accuracy) and via a binary classification model (90% accuracy).

6.4 BERT - Glazkova (2021)

This paper analyses BERT (as well as modified BERT models such as RoBERTa) and how they can be used for text classification. The data used in this paper is a set of scientific papers that are classified into 7 different categories.

The paper shows that using a Feedforward Neural Network (FNN) on top of BERT can achieve a 91.76% accuracy on the dataset.

7 Timeline

8 Risk Management

There are several factors that pose a risk to this project. Below is a table to illustrate what risks are prevalent, how big of an impact they will have, and how to counteract these risks.
The scores are ranked from 1-5. 5=high, 1=low.

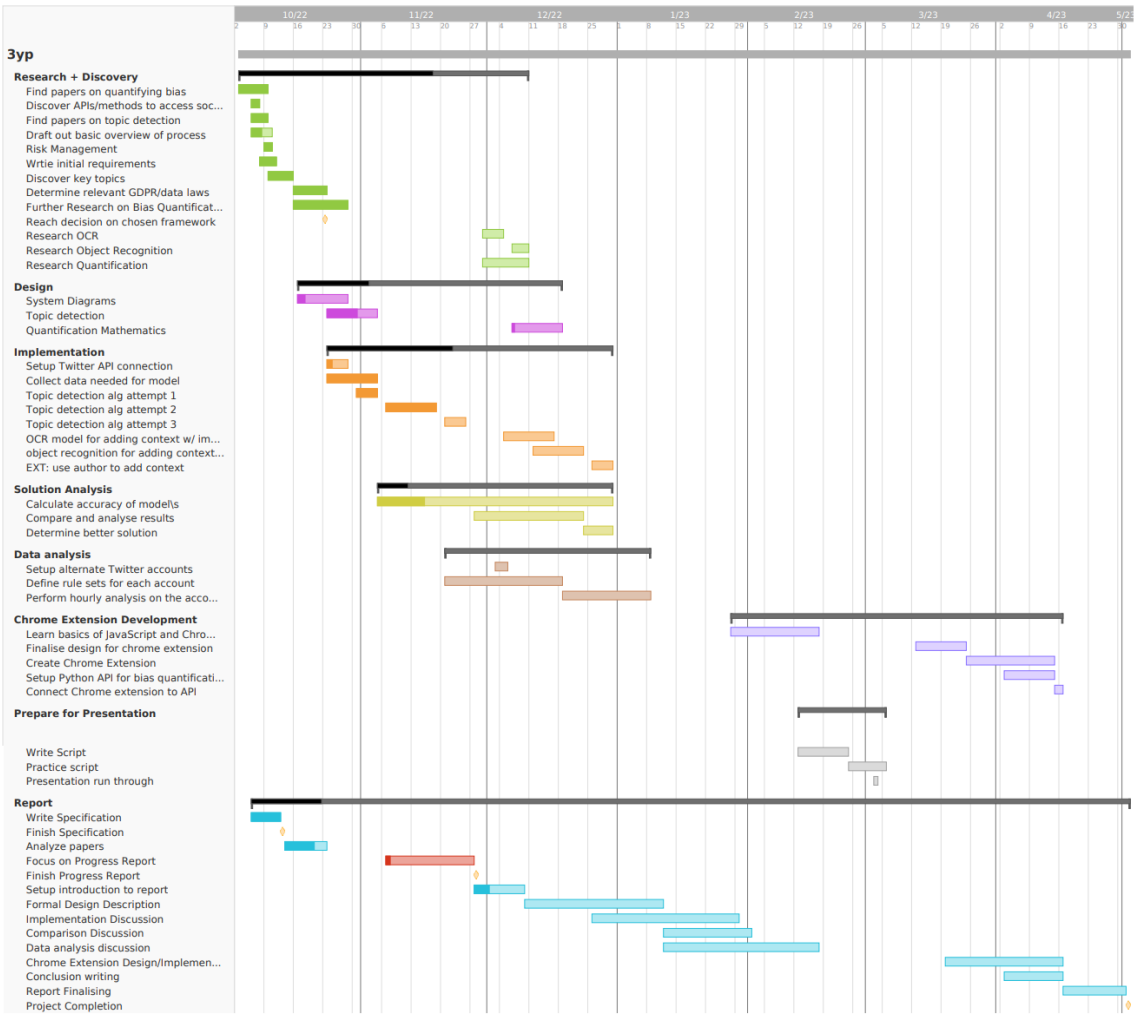

Figure 2: Project Timeline

9 Ethical/Legal Considerations

As mentioned in Section 3.3 when storing information of users - such as demographics, identifiable information, and social media usage - It is important to ensure we follow relevant GDPR and data protection laws.
When gathering data, we will also required volunteers. The volunteers will need to have a thorough understanding of what information we are gathering and we need to ensure we use this data lawfully.

References

Glazkova, Anna. Identifying topics of scientific articles with bert-based approaches and topic modeling. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 98–105. Springer, 2021.

Husby, Stephanie & Barbosa, Denilson. Topic classification of blog posts using distant supervision. In *Proceedings of the Workshop on Semantic Analysis in Social Media*, pages 28–36, 2012.

Instagram, . How Instagram determines which posts appear as suggested posts | Instagram Help Centre. URL `https://help.instagram.com/381638392275939`.

Litou, Iouliana & Kalogeraki, Vana. Pythia: A system for online topic discovery of social media posts. In *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, pages 2497–2500, 2017. doi: 10.1109/ICDCS.2017.289.

Peng, Xian & Han, Chengyang & Ouyang, Fan & Liu, Zhi. Topic tracking model for analyzing student-generated posts in spoc discussion forums. *International Journal of Educational Technology in Higher Education*, 17(1):35, Sep 2020. ISSN 2365-9440. doi: 10.1186/s41239-020-00211-4. URL `https://doi.org/10.1186/s41239-020-00211-4`.

Shin, Donghee. How do users interact with algorithm recommender systems? the interaction of users, algorithms, and performance. *Computers in Human Behavior*, 109:106344, 2020. ISSN 0747-5632. doi: https://doi.org/10.1016/j.chb.2020.106344. URL `https://www.sciencedirect.com/science/article/pii/S0747563220300984`.

Snow, Rion & Jurafsky, Daniel & Ng, Andrew. Learning syntactic patterns for automatic hypernym discovery. *Advances in neural information processing systems*, 17, 2004.

Risk	Likelihood of Occurring	Impact to Project	Impact Mitigation	New Impact
Personal issue causing delay in schedule	4	3	I have purposefully planned my project with room to	1
Chosen design philosophy fails to produce useful bias metrics	3	4	Firstly, this project is mainly research based, so no matter the outcome we will gain something from the comparison between approaches. For the Software side of the project we could instead use an already known method if ours does not work.	2
Changes to Twitter API	2	1	Any changes to the API should not dramatically affect this project. At worse it could involve some minor code refactoring to correct endpoints/REST queries.	1
Laptop dies	1	5	Using git+github for version control throughout this project (for code and report), no matter what happens to my own laptop, the codebase will be accessible from any machine	1
Unable to recreate other papers methods	4	4	Instead, I can take the papers results as true (after analysis of results credibility). I can also just create the chrome extension using my own method	1
inability to gather enough users to generate data for analysis	2	4	Can perform analysis based on other metrics than demographics (e.g. who they follow, what the like, etc.) This data can be generated artificially	2
Twitter website/API being taken off the internet	2	5	This is a very unlikely scenario, but if it were to happen, I would have to find another source of data to analyse.	1

Table 1: Possible risks