# Using NLP Analysis to Categorise Statements to Values on the Political Compass

Alexander Martin Hey
Student Number: 170466662
Supervisor: Arkaitz Zubiaga
*Student of Queen Mary University Of London*
London, United Kingdom
Student Email: a.m.hey@se22.qmul.ac.uk
Email: alex@alexhey.co.uk
Computer Science MSc

***Abstract*—This paper intends to address the challenges posed by classification in online political discourse, particularly on social media platforms. By utilising state-of-the-art NLP techniques and a standardised political stance metric, the study aims to develop a more nuanced classification model for understanding individual political viewpoints. This will be done using the analysis of the newly obtained dataset and begin the exploration of multiple classification methods. The research seeks to offer a complete approach to addressing the complexities of political opinions in the digital age, ultimately contributing to a better understanding of political standing within social media environments.**

## I. Introduction

Social media has been increasingly used as a platform for online political discourse in recent years. Both users and companies have an underlying bias that is often not directly disclosed, making seemingly neutral speech hold underlying bias. In addition to this, opinions can also be misleading if only classified on how left or right a view is. This political bias can be seen in Muhammad Ali's 2021 Paper Ali et al. (2021) where it was determined that 'Facebook is wielding significant power over political discourse through its ad delivery algorithms'. This can also be proven by Daniel Sussers' analysis of their 2019 paper Susser et al. (2019). In section II.C 'Psychographic Profiling and Election Influence', it is summarised that it is possible to obtain 'gender, sexual orientation, race, religion, political views, relationship status, substance use, and size and density of friendship networks' by utilising a user's digital signature.

Osemar A. Caetano's paper 'Using sentiment analysis to define Twitter political users' Caetano et al. (2018) does this to some extent by classifying support of Trump or Hillary in the 2016 election. This attempt worked well however missed the multi-faceted nature of political opinions. Another example that has addressed this issue before is Fabian Falck's 'Sentiment political compass: a data-driven analysis of online newspapers regarding political orientation' Falck et al. (2018). This paper manages to begin addressing the problem of political classification by utilising the political compass to plot individual statements of news organisations. This is a very good methodology to use and would be worth expanding in this report. Falck's model focused heavily on German news organisations so more effort would be needed to make classification work on more general statements.

To further advance this initial research, this project intends to use a more standardised way of political standing classification technique; The Political Compass Organisation (n.d.). The Political Compass is a commonly used political stance metric that has been used consistently for many years Google (2023). The core principle is plotting data on two separate axes; economic and social Compass (2017). This allows for a more nuanced description of any individual political viewpoint.

This project will attempt to achieve this goal by utilising state-of-the-art NLP techniques to create an NLP classification model based on a newly created dataset. The data will be collected from multiple different sources and will be manipulated in ways that allow this model to learn in the best way possible. Then in addition to this, I intend to use the now prominent ChatGPT3.5 API to manipulate the original data set into one more tuned towards human communication rather than the raw format.

The key components of this work can be followed through from the roots. I have analysed existing reports in order to see where previous research has been done in this sector and intend to replicate/improve on some of their findings. In addition to this, the generation of a new dataset to complement the NLP model should be open and available for future use. I also intend to use multiple methods to act as classifiers and will be analysing the potential use cases and challenges of these approaches.

This research is important as it allows a more complex approach to simple classification that is often done by people when judging someone's political standing. A more broad use system instead of the current focussed ones may prove to be more useful in a social media environment.

The aims of this project are to answer the following questions.

1) Can open data be used to build a large dataset of political cases in a non-biased way?
2) Can this data be used to train an NLP model to correctly classify records in the dataset and to what level of

accuracy can this be done?
3) Can the same dataset be used on an individual's public speech?

This will be done by achieving the following objectives:

1) Analyse the current political compass scale in order to create a new dataset with political classification labels.
2) Build and train up 2 different models using different methodologies.
3) Utilise cross-validation to confirm and test the results.

## II. Non-Bias Conditions

Due to the nature of this report, the author must acknowledge that he holds his own biases and beliefs. While this is done on a personal level, the author will take all necessary steps to avoid biases in this report. This will be done by the utilisation of the following rules:

1) All data must be open and accessible to the public and sourced thoroughly to maintain data integrity.
2) A full and detailed attempt must be made to find self-identified bias sources.
3) In the case that a self-identified source cannot be identified, at least 2 supporting sources must be cited to confirm categorisation.

To ensure that I followed each of these conditions, I have had multiple people from all wakes of politics review my work and any assumptions that I have had to make. In the cases where any bias was seen, I documented this and worked on a perceived solution using the help of the reviewer. This solution was then re-reviewed by the entire panel to confirm that the solution also obeyed the above rules. Cases, where this happened, will be documented in the relevant sections.

## III. Related Work

Before starting this project, it was important to analyse previous work that had been completed in a similar sphere of knowledge. For this purpose, I wrote the following questions to research.

- Has any research been done around this area before?
- Have any issues in this area been encountered before?
- If implementations exist with similar problems, what methods of NLP were used?

The first paper of interest was Josemar A. Caetano's paper 'Using sentiment analysis to define twitter political users' classes and their homophily during the 2016 American presidential election' Caetano et al. (2018). This project was to use Sentiment Analysis to categorise any tweet into one of 6 categories comprising candidate support and sentiment regarding the tweet. This paper was instrumental in the creation of this project as it gave a useful way to filter and pre-process similar data to what could be provided through Twitter. The report however differs significantly from this project's end goal as it would only categorise into one category and would not offer a figure of how strong the belief is.

Another paper of use was Anna Stavrianou's 'NLP-based Feature Extraction for Automated Tweet Classification' Stavrianou et al. (2014). This paper used multiple feature extraction techniques to provide a similar classification to tweet structure. This paper used feature extraction in many forms as a pre-processing technique and achieved reasonable results. For this project, BERT encoding will be used to contrast against this study's results.

Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis by Lei Zhang Zhang et al. (2011) also attempted to achieve a similar goal but on a broader scale. According to the abstract, 'Twitter's unique characteristics give rise to new problems for current sentiment analysis methods' Zhang et al. (2011). This would prove to be a problem with non-BERT model learning in this project however some of the issues such as problematic recall and F1 Score should be replicated when attempting LinearSVM.

Fabian Falck has also written multiple papers regarding text classification. His classification output provided fits a similar format as to what this project aims to achieve. In his 2020 paper 'Measuring Proximity Between Newspapers and Political Parties: The Sentiment Political Compass' Falck et al. (2020) a new compass was coined by Falck called the 'Sentiment Political Compass (SPC)'. This seems to be based on the original political compass however is currently unavailable at the advertised link. In addition to this, another paper was released 'Sentiment political compass: a data-driven analysis of online newspapers regarding political orientation' Falck et al. (2018) which attempts to complete the same task as this project but with a substantially different dataset; newspaper articles. The paper managed to utilise a model that allowed the creation of political compass positions using standing news organisations' data. This was achieved using relative sentiment to prominent political candidates. This was an initial consideration regarding data collection for this project however the organisations categorised towards a German audience. This resulted in a problem that classification would become more tailored to a specific subset of people causing it to contrast the project goals. As such, openly biased news organisations were chosen.

Finally was Wen Chen's paper 'Neutral bots probe political bias on social media' Chen et al. (2021). Whilst this paper confirmed that there was indeed bias that could be determined from users' speech on social media platforms, it did identify something that I want to avoid in this paper. When defining 'low-credibility' outlets, they refer to a 3rd party fact-checking organisation to classify these groups. Whilst this works in their paper, I will only be using 1st party data unless none is obtainable. In this case, 2 supporting sources will be used to confirm the prognosis.

### A. Answering Research Questions

From the related reading, I can now confidently answer all of the stated research questions. To begin with, previous research in this area has been done however there is still significant means for expansion on existing methodologies. Research spans significantly in this sphere from the implementation side of things to the theoretical.

To answer the second question, Lei Zhangs Zhang et al. (2011) identified multiple issues regarding political classification on some social media structured data resulting in worse F1 and Recall scores. It is also noted that Falck created the 'Sentiment Political Compass' as a new scale as opposed to my planned method which is using the official Political Compass. There are many reasons for this as will be discussed in the Political Compass Analysis section of this report.

Finally, it is possible to see that sentiment analysis is a key area of research among these related papers. This area of NLP seems to be thoroughly explored and so it would be beneficial to attempt different methodologies.

### B. Uniqueness Factor

After analysing the above paper. I believe that this project differs from the existing research in many ways. Firstly, almost all existing studies do not attempt to use a standardised scale. This makes repeating the study substantially more difficult as the current methods are not attuned for general use but are instead made for these specific use cases. This project intends to utilise a commonly used scale (The Official Political Compass) and work in a much more general way than previous papers.

Secondly, the dataset that I will be utilising will be generated based on a new dataset with self-defined data rather than referring to a 3rd party organisation as was done for 'low-credibility content' in Chen et al. (2021). In addition, this dataset will be parsed through a pre-existing large language model to see if it is feasible.

Lastly, this paper will not be primarily using sentiment analysis and instead will focus on other methods of classification. This will be LinearSVM and BERT Classification as they range in complexity and ease of implementation. BERT will be implemented specifically as it is a very recent implementation that has yet to be tested in this context.

## IV. Initial Plan of Implementation

The initial plan for the project was broken into 2 different layers. This section will detail the integration layer where data is collected from multiple online sources and parsed through into an SQL Database. This layer focuses on gathering data from various political categories, such as Libertarian, Left, Right, and Authoritarian, by employing different techniques and tools. The data collection process involves parsing HTML, utilising APIs, and employing web emulators to extract the required information from sources like the Cato Institute, Occupy Democrats, National File, and Al Jazeera. This section of the project aims to ensure a comprehensive and diverse dataset for further analysis and classification.

The second step of this implementation is the classification step. This entails all faucets' NLP and ML as well as the output classification. Figure 1 shows the implementation of the BERT model, which takes the information from the SQL database and parses it through the TensorFlow API to first preprocess the text, and then encode the text. Once this is done, a model that consists of multiple layers as shown in Fig. 2 was constructed.

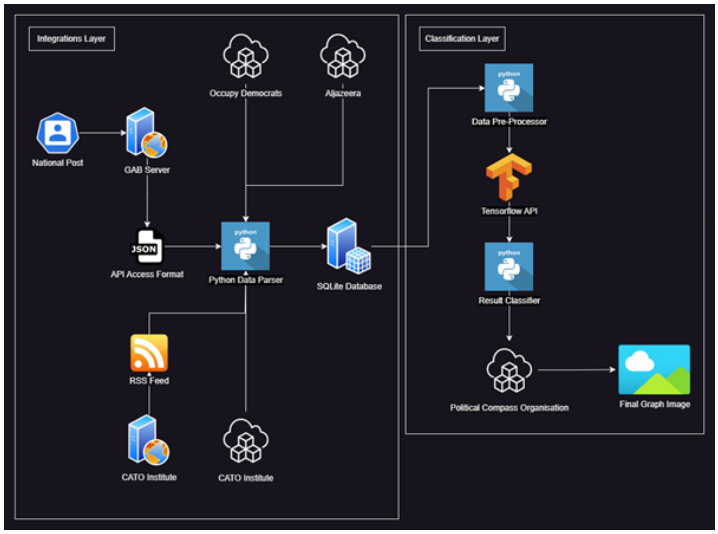


Fig. 1. Diagram of Initial Plan of Implementation.

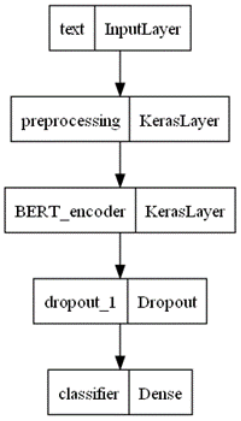


Fig. 2. NLP Bert Model Layout using Keras utility plot_model.

## V. Tools Used

To implement the full process, a wide suite of tools was utilised. As this was almost exclusively implemented in Python, I made use of the following library (requests.txt). The related table can be seen under the Appendices section in Table XIII.

## VI. Political Compass Analysis

The political compass is a standardised framework used to categorise political ideologies and positions. It provides a two-dimensional model that categorises political stances on 2 different axes: the economic left/right on the X axis and social authoritarian/libertarian on the Y axis. Using these 2 axes allows a more faceted analysis of political opinions and parties.

### A. Criticisms

The creator of the site and the company that runs the organisation have attempted to remain anonymous for the duration of the website's uptime. This means that it is nearly

impossible to see the find if the creators have any sort of bias that the questions may push.

In addition to this, PACE NEWS LTD (The copyright holders) have not made clear how the test is scored or how each question has an effect on your final result. This was criticised by Tom Utley in 2001 Utley (2001) especially as the site has historically tried to plot prominent political figures using their public statements (Figure 3). This is almost impossible to do in an unbiased way as the scale that each question is judged on is from 0-5 (Strongly Disagree to Strongly Agree).

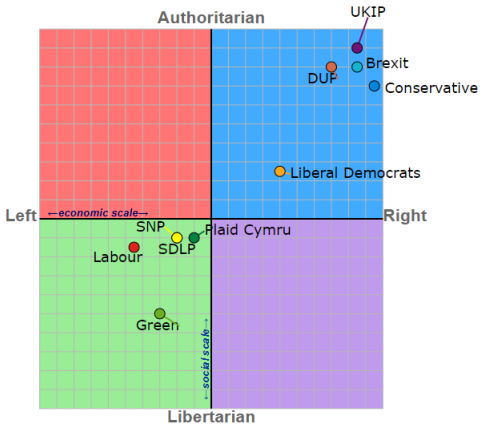


Fig. 3. Political Organisation Plot from the official source for UK Parties in 2019.

## VII. Data Acquisition

To allow the creation of a model that could correctly classify positions on the political compass, it was important to collect a relevant dataset with correctly labelled categories. The way that each axis is scored is on a scale of -10 to 10. For example, if someone is a (5,5) you would be seen as someone on the economic right whilst also being socially authoritarian. Due to the complexity of this scoring system, the author created a plan to estimate correct scoring.

Instead of generating a dataset of labelled individual points, data can be labelled by one of the relevant 4 categories (left, right, authoritarian, libertarian). From here the data can be split down into 2 separate datasets; x-axis (left-right) and y-axis (authoritarian, libertarian). The model then only has to predict which category the statement fits into on each axis. From this point, we can use the confidence of prediction on each axis to scale the output to lay on the line from -10 to 10.

### A. Calculations

Here are the formulas used to calculate values from probability dictionaries returned by both classifiers.

*1) X Axis Calculation:*

$$M = ProbabilityDictionary \tag{1}$$

$$T = MAX(M) \tag{2}$$

$$P_{Left} = M[_{left}] \tag{3}$$

$$P_{Right} = M[_{right}] \tag{4}$$

$$\delta = \begin{cases} -10, & \text{if } P_{Left} > P_{Right} \\ 10, & \text{if } P_{Left} < P_{Right} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

$$x = ((T - 0.5) * 2) * \delta \tag{6}$$

*2) Y Axis Calculation:*

$$M = ProbabilityDictionary \tag{7}$$

$$T = MAX(M) \tag{8}$$

$$P_{Libertarian} = M[_{Libertarian}] \tag{9}$$

$$P_{Authoritarian} = M[_{Authoritarian}] \tag{10}$$

$$\delta = \begin{cases} -10, & \text{if } P_{Libertarian} > P_{Authoritarian} \\ 10, & \text{if } P_{Libertarian} < P_{Authoritarian} \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

$$Y = ((T - 0.5) * 2) * \delta \tag{12}$$

### B. Category Breakdown

To generate the dataset combinations of acquisition techniques would be used to build a dataset. The following is a breakdown of each category (left, right, authoritarian, libertarian) and how the data was collected.

*1) Left:* For this source, I opted for Occupy Democrats. This is an openly partisan party as stated on the footer of every page on their website; “Occupy Democrats is a political organization and information website that provides a new counterbalance to the Republican Tea Party” Democrats (2023). As this is a company that is prevalent on social media, this was a prime source.

To acquire this data, the requests library and BeautifulSoup were used to obtain this data from the HTML source code. From here, headlines were taken and logged in the SQL database.

*2) Right:* National File was the primary source for right-leaning news. Their partisanship is stated by the following comment on their about page ‘National File’s team of writers includes veterans of prominent conservative outlets including Breitbart and The Daily Caller’ File (2023).

Due to their heavy social media presence on the site Gab Gab (2023), this seemed like a prime choice. For the data collection here, an inbuilt API was utilised to collect headlines and tags from all posts from the related account.

*3) Libertarian:* The Cato Institute is an openly Libertarian organisation stating that '... the Cato Institute is to create free, open, and civil societies founded on libertarian principles' Institute (n.d.). They have had a blog with frequent posts since 2006.

This site has an open RSS Feed which was used to collect data as required. As this source is more formalised and does not have as large a social media presence, I expect the results of classification to be worse than the equivalent X-axis predictions.

*4) Authoritarian:* This source was much more difficult to ascertain due to the nature of authoritarianism. From extensive research, I could not find an organisation that explicitly claims to be authoritarian. To work around this problem. I referred to multiple democratic index sources including the EIU (Economist Intelligence Units) Democracy Index 2022 Unit (n.d.) and the V-Dem's Democracy Report V.-Dem (n.d.) which both point towards progressively more authoritarian countries.

Initially, it was planned to use a news source as near to the bottom of the list as possible (sorted by least to most authoritarian) however this caused an issue in a later step as the more extreme examples are either not present on the surface web or are overly tailored to the nation's problems and issues. For example, China is classed as Authoritarian by EIU Unit (n.d.) however one of the prominent news sources is Xinhua Xinhua (n.d.) has a few different issues associated with it.

Firstly, the site that is directed to by the host is an English-tailored version meaning that news may be filtered or altered causing non-partisan news to be included. In addition to this, most of the news is focused on China itself resulting in problems when training the model. This caused the model to believe anything that mentions China to be authoritarian.

The final decided news source was Al Jazeera as it is run by Al Jazeera Media Network Jazeera (n.d.). This was written and acknowledged in Tal Samuel-Azran's critique 'Al-Jazeera and Qatar's soft power' Samuel-Azran (n.d.). This source would be much more appropriate as while not as strongly authoritarian, the news that it reports is global and therefore better to train the data off of.

## VIII. Classification Methods

To generate an NLP model for classifying statements, I used 2 different main methodologies. These went through repeated testing and improvements of which the whole log can be found on GitHub Hey (2023). The methods that were used are broken down in the following sections.

### A. Linear SVC Classification

*1) Pre-processing:* Pre-processing occurs on each string with the following process:

- The entire sentence is converted to lowercase.
- Special characters are filtered out.
- Stopwords are filtered using Natural Language Toolkit (nltk).
- Lemmatisation of tokens (nltk).
- Manual word filtering.

This pre-processing was done in an effort to simplify the learning process. However, it may have caused some unforeseen issues as a small amount of context was lost in this process. An additional step I also implemented was POS tagging. This had the unforeseen consequence of significantly increasing learning time and so it was removed. A similar methodology was used in Stavrianou et al. (2014) paper which resulted in a lower f1 and recall score so it would potentially explain future results in this project. In addition to this, due to the overall data collection process, I would expect authoritarian learning results to be slightly lower than originally predicted.

*2) Training:* Before starting training, the data was split into 4 categories:

- X-Axis Data – (Left/Right Records) 80:20 train:test data
- Y-Axis Data - (Libertarian/Authoritarian Records) 80:20 train:test data

In addition to this, the data was balanced so that the split between each category was equal. This resulted in each category having an even 25% split of the dataset. When splitting this again into its X/Y sets, this would result in an even 50:50 split.

For LinearSVM, the majority of the time was spent pre-processing the data due to the amount of filtering and tokenisation that was required. However, this resulted in a well-learned set that performed well against the training data.

### B. BERT Classification

A more complex methodology that seemed appropriate for use in this task is BERT (Bidirectional Encoder Representations from Transformers) Devlin et al. (2018). After analysis, the BERT methodology learns trends and classes by building up connections between words in a string by utilising a pre-trained model. This way, each word used already has context prescribed by a global model in addition to then utilising a dense layer to calculate classifications for a unique category. This method was not used in any of the related works due to its relatively recent development. If example problems are to be believed, this model could provide significant benefits over LinearSVM. It is important to note that this method of classification has not been implemented or tried in any of the papers cited in the related work and therefore may be used to improve existing methodologies.

*1) Pre-processing:* Pre-processing using BERT is split into 2 separate categories, the preprocessing step and the encoding step. Both of these are created as layers to the network and are done by interfacing with Tensorflow's Pre-Trained BERT models. After experimentation, bert_en_uncased_L-4_H-512_A-8/1 TensorFlow (n.d.*a*) seemed to be the best choice as it is a small enough dataset that encoding does not take an unfeasible amount of time however good connections can still be made between strings of words. In addition, I also used the corresponding BERT pre-processor to go with this encoder as designated by the TensorFlow documentation. TensorFlow (n.d.*b*). Secondly, the same efforts were put in place as with

LinearSVM to split and balance the data so that there is no partisanship in data sizes.

*2) Training:* As shown in Figure 2, additional layers were added to the model to allow for better training. The first additional one was a drop-out layer. This was put in place to avoid overfitting the model to the original data. The second was a dense layer responsible for outputting the results of the network. The model was trained with multiple parameters however the ones that success was found in were the following:

- Epochs – 5
- Learning Rate – 3e-5
- Batch Size - 174 (Dynamically Calculated)

## IX. Results

To calculate the results for both models attempted, cross-validation was used on both axes. This allowed for testing against known data points. This however does pose a problem regarding extremities on the political compass. As each sample will be classified into 2 of 4 total categories, it is important to note that it is impossible to check the accuracy of the compass against each individual statement. This is due to the fact that the initial dataset did not have positions on the compass recorded as no such data set exists. To resolve this, the calculations detailed in section VII-A were used to convert the certainty matrix to positional coordinates.

For the sake of evaluation, results and accuracy figures will only based on correct quadrant classification as that is the data that the model is trained on. In addition to this, depending on the accuracy, manual tests will be done using individuals and statements on a more practical level to confirm severity.

### A. LinearSVM

Due to LinearSVM's overall simplicity, these results are reasonable but significantly improvable. Here are the X-Axis results.

TABLE I
X-Axis Train Results

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| **Left** | 0.53 | 0.99 | 0.69 | 646 |
| **Right** | 0.90 | 0.11 | 0.19 | 646 |

TABLE II
X-Test Train Results

| | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|
| **Overall** | 0.71 | 0.54 | 0.43 | 0.54 |

Overall on the X-Axis, the model is very susceptible to classifying false positives, especially when the text is left-leaning. In addition to this, the right classification suffers the opposite problem with an inadequate F1 and Precision score. This is similar to what was predicted and already seen in Stavrianou et al. (2014) paper regarding feature extraction. More than likely, this complex task is too complex for a simple algorithmic-based solution such as linear SVM. This being demonstrated though shows that there is still the ability to correctly classify data on the X-Axis over 50% of the time.

TABLE III
Y-Axis Train Results

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| **Authoritarian** | 0.70 | 0.97 | 0.81 | 647 |
| **Libertarian** | 0.94 | 0.58 | 0.71 | 646 |

TABLE IV
Y-Test Test Results

| | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|
| **Overall** | 0.82 | 0.77 | 0.76 | 0.77 |

The Y-Axis LinearSVM classifier was much more successful than the X-Axis one due to its overall better figures. The F1 Scores and recall for both Libertarian and Authoritarian are within very good ranges and so it is therefore a much better classifier. My assumption for why this is the case is due to the complex nature of the Cato Institute's article formatting. I believe that it is uniquely identifying structurally causing the model to predict more reliably by analysing sentence format rather than content. This proved to be almost a perfect axis classification.

### B. BERT Classification

TABLE V
BERT X-Axis Model Training Statistics

| Epoch | Loss | Accuracy | Precision | Recall |
|---|---|---|---|---|
| **1** | 0.6907 | 0.5647 | 0.6023 | 0.3814 |
| **2** | 0.4014 | 0.8033 | 0.8969 | 0.6855 |
| **3** | 0.2249 | 0.9057 | 0.9423 | 0.8643 |
| **4** | 0.1671 | 0.9325 | 0.9588 | 0.9039 |
| **5** | 0.1341 | 0.9491 | 0.9744 | 0.9313 |

TABLE VI
BERT X-Axis Model Test Statistics

| Loss | Accuracy | Precision | Recall |
|---|---|---|---|
| 0.1148 | 0.9601 | 0.9744 | 0.9449 |

TABLE VII
BERT Y-Axis Model Training Statistics

| Epoch | Loss | Accuracy | Precision | Recall |
|---|---|---|---|---|
| **1** | 0.6747 | 0.6652 | 0.7321 | 0.5208 |
| **2** | 0.3981 | 0.8136 | 0.9055 | 0.7001 |
| **3** | 0.2080 | 0.9163 | 0.9412 | 0.8880 |
| **4** | 0.1457 | 0.9455 | 0.9558 | 0.9341 |
| **5** | 0.1110 | 0.9581 | 0.9618 | 0.9539 |

TABLE VIII
BERT Y-Axis Model Test Statistics

| Loss | Accuracy | Precision | Recall |
|---|---|---|---|
| 0.1194 | 0.9563 | 0.9659 | 0.9461 |

From this, you can see that the accuracy and loss of this method were heavily improved over the LinearSVM model with over 90% of categorisation accuracy. To add to this, the number of Epochs this model was trained over was relatively small and loss may not have plateaued yet.

In addition to this, the result of this method seems to be significantly higher than most NLP papers previously cited in the report. This shows that the use of pre-trained models is very effective when regarding political position classification.

One of the problems that occurs when using the current political compass is that it is almost impossible to test in a real-world situation due to its inherent subjectivity. In future studies, more standardised scales such as the one used in Falcks' Falck et al. (2020) piece could be used.

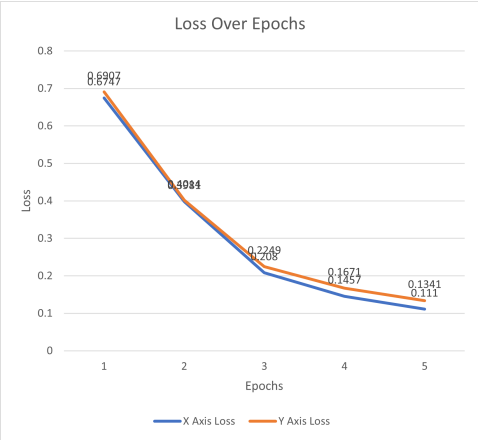


Fig. 4. BERT Loss Graph.

From 4, you can see that with a learning rate of 3e-5, over 5 epochs, the loss begins to plateau to near 0 as expected. One thing that is difficult to determine with this breakdown is if the graph would continue to decrease further or if it has reached the minimum value.

## X. Discussion and Improvements

Whilst the results from both models were reasonably successful, when putting user-made examples through, sometimes the models would categorise wildly incorrect answers. Whilst this is a rare occurrence, the model requires headline adjacent formatting data to correctly identify these categories. This was not the original plan of this report and so improvement was required here.

After looking at the results, it is possible to discern that the model has learnt the formatting of categorical data rather than the content of the statements. I believe that will be resolved once the new OpenAI tweet formatted dataset has been created and used for training.

In addition to this, due to the way that the coordinate positions are calculated, the BERT prediction certainty causes results to appear on the extreme boundaries of the compass. This is also closely related to another issue with the current classification methods; Neutral Statement classification. This issue occurs when a non-political statement is put through the current models that are only trained to classify 2 of 4 classes. I originally expected neutral statements to come back as roughly 50% Left Leaning and 50% Right Leaning on the X-axis for example. If this was the case, then the calculation that was put in place would correctly classify to roughly (0,0). In practice, this does not seem to be the case.

In a future study, it may be possible to resolve both of these issues by utilising a 3rd classifier. This model would be able to determine if the data is political or neutral and use that factor between 0 and 1 as a scalar to plot neutrality. Unfortunately, this was not possible to complete in this report as it requires the obtaining of a brand new dataset of both neutral and political tweets as well as the training and implementation of a new model. The extent of this problem can be seen in a report published by Pew Research Centers Bestvater et al. (2022) where it is said that only '33%' of Twitter traffic is political.

## XI. Improvements to the Initial Methodologies

After reviewing the results from the BERT model, I was relatively happy with the classification success rate however some significant improvements could be made. To begin with, the dataset carried news headlines from partisan sources. Due to the nature of the subject matter, any actual categorised data would be nearly impossible to obtain and so an additional method was implemented.

### A. OpenAI

To improve the formatting of the original dataset, OpenAI's ChatGPT3.5 API OpenAI (2023) was utilised to convert each of the headlines in the dataset to 'tweet' versions. This was done using the following prompts:

- The statement following the next occurrence of the delimiter '—' is a news headline that ends at the following '—' delimiter. The final word of the prompt will correspond to a political category. Using these 2 pieces of information you should convert the news headline into a tweet as if it was written by an average Twitter user.",
- "The tweet must be in support of the category ideology. Libertarians uphold the belief in liberty and personal freedoms. Authoritarians should agree with central power to preserve the political status quo (often done through enforcement or military action) along with an anti-freedom mindset. Examples of these are dictatorships. This group is usually very nationalistic and is pro-interference",
- "The following people are prominent political figures followed by their standing. Use this context to base the tweets: Joe Biden - Left, Donald Trump - Right, Barack Obama - Left, George W. Bush - Right, Bill Clinton - Left, Hillary Clinton - Left, Mike Pence - Right, Ron DeSantis - Right, Gavin Newsom - Left. You should avoid using tags in the tweet and must not express the category in plain text, only imply the users' beliefs."

Every record in the original data set was then run through this prompt throughout a period of 12 hours and then saved to a new database.

Table XVI in the appendices section is an example of 2 of the records and the conversion between the original statement and the OpenAI tweet format.

From here, a new BERT model was trained using the original methodology. This was done with a few minor changes to the learning parameters.

- Epochs – 10
- Learning Rate – 3e-5
- Dataset – OpenAI tweets

### B. Updated Results

TABLE IX
IMPROVED BERT X-AXIS MODEL TRAINING STATISTICS

| Epoch | Loss | Accuracy | Precision | Recall |
|---|---|---|---|---|
| **1** | 0.7793 | 0.4895 | 0.4734 | 0.1873 |
| **2** | 0.7080 | 0.5289 | 0.5669 | 0.2446 |
| **3** | 0.6245 | 0.6081 | 0.7066 | 0.3696 |
| **4** | 0.5413 | 0.6974 | 0.7633 | 0.5722 |
| **5** | 0.4720 | 0.7554 | 0.8058 | 0.6729 |
| **6** | 0.4276 | 0.7853 | 0.8345 | 0.7116 |
| **7** | 0.3551 | 0.8096 | 0.8473 | 0.7554 |
| **8** | 0.2935 | 0.8370 | 0.8731 | 0.7888 |
| **9** | 0.2157 | 0.8573 | 0.8873 | 0.8185 |
| **10** | 0.1126 | 0.8709 | 0.8929 | 0.8428 |

TABLE X
IMPROVED BERT X-AXIS MODEL TEST STATISTICS

| Loss | Accuracy | Precision | Recall |
|---|---|---|---|
| 0.2401 | 0.8354 | 0.8575 | 0.8045 |

TABLE XI
IMPROVED BERT Y-AXIS MODEL TRAINING STATISTICS

| Epoch | Loss | Accuracy | Precision | Recall |
|---|---|---|---|---|
| **1** | 0.8154 | 0.6029 | 0.5896 | 0.6773 |
| **2** | 0.5811 | 0.6656 | 0.8012 | 0.4406 |
| **3** | 0.3445 | 0.8684 | 0.9318 | 0.7950 |
| **4** | 0.2157 | 0.9208 | 0.9435 | 0.8952 |
| **5** | 0.1855 | 0.9299 | 0.9427 | 0.9154 |
| **6** | 0.1647 | 0.9389 | 0.9497 | 0.9270 |
| **7** | 0.1500 | 0.9462 | 0.9546 | 0.9369 |
| **8** | 0.1280 | 0.9542 | 0.9669 | 0.9406 |
| **9** | 0.1126 | 0.9594 | 0.9728 | 0.9451 |
| **1** | 0.0972 | 0.9651 | 0.9708 | 0.9592 |

TABLE XII
IMPROVED BERT Y-AXIS MODEL TEST STATISTICS

| Loss | Accuracy | Precision | Recall |
|---|---|---|---|
| 0.1924 | 0.9455 | 0.9478 | 0.9431 |

From Tables IX, X, XI, XII, you can see that the accuracy of the improved model increased significantly to an average of 90% on each axis. In addition to this, the model fairs much better against real-world statements.

## XII. CONCLUSION

This project aimed to use multiple NLP techniques to classify statements onto a political compass. By using LinearSVM and BERT the project achieved impressive results in accurately classifying political statements. With the help of OpenAI's ChatGPT 3.5, it proved that it can be used to generate a usable data set when given the correct prompts.

This report details the ways and methods that were utilised to fully achieve the aims and objectives of this project. In addition to this, further avenues of investigation and potential improvements were also discovered and have also been detailed to help improve further work in the field of NLP political classification.

Whilst this project has been a significant success regarding political stance classification, there are still problems that need to be addressed in a future project. Dealing with statements lacking context or which do not have a political nature should be a primary goal. According to a report published by Pew Research Centers Bestvater et al. (2022), '33%' of Twitter traffic is political meaning that in the majority of cases, the model should give a Neutral position. An additional issue that still needs to be addressed is how the scaling calculation works. Currently, if the model is certain of a prediction to any one class, the extremity of the position is shown much greater than should be, this should be addressed by conducting a study of people so that individual statements can be given political compass values.

Overall, this project has laid a strong foundation for the use of NLP in political classification online, offering a model that can deal with a wide variety of statements. With NLP progressing at the current rate, better classification techniques could allow for this concept to enable a deeper understanding and transparency when interpreting views in the ever-evolving social media space.

## XIII. ACKNOWLEDGEMENTS

Many thanks to Dr. Arkaitz Zubiaga for his invaluable guidance and expertise throughout this implementation project. His insights were crucial to the success of this project and it would not have been possible without them.

## REFERENCES

Ali, M., Sapiezynski, P., Korolova, A., Mislove, A. & Rieke, A. (2021), Ad delivery algorithms: The hidden arbiters of political messaging, *in* 'Proceedings of the 14th ACM International Conference on Web Search and Data Mining', pp. 13–21.

Bestvater, S., Shah, S., River, G. & Smith, A. (2022), 'Politics on twitter: One-third of tweets from us adults are political'.

Caetano, J. A., Lima, H. S., Santos, M. F. & Marques-Neto, H. T. (2018), 'Using sentiment analysis to define twitter

political users' classes and their homophily during the 2016 american presidential election', *Journal of Internet Services and Applications* **9**(1).

Chen, W., Pacheco, D., Yang, K.-C. & Menczer, F. (2021), 'Neutral bots probe political bias on social media', *Nature communications* **12**(1), 5580.

Compass, T. P. (2017), YouTube.
**URL:** *https://www.youtube.com/watch?v=5u3UCz0TM5Q*

Democrats, O. (2023), 'Occupy democrats - news'. Retrieved from Occupy Democrats:.
**URL:** *https://occupydemocrats.com/category/news/*

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2018), 'Bert: Pre-training of deep bidirectional transformers for language understanding', *arXiv preprint arXiv:1810.04805* .

Falck, F., Marstaller, J., Stoehr, N., Maucher, S., Ren, J., Thalhammer, A., Rettinger, A. & Studer, R. (2018), Sentiment political compass: a data-driven analysis of online newspapers regarding political orientation, *in* 'The Internet, Policy & Politics Conference', number 3.

Falck, F., Marstaller, J., Stoehr, N., Maucher, S., Ren, J., Thalhammer, A., Rettinger, A. & Studer, R. (2020), 'Measuring proximity between newspapers and political parties: the sentiment political compass', *Policy & internet* **12**(3), 367–399.

File, N. (2023), 'About page'. Retrieved from.
**URL:** *https://nationalfile.com/about/:*

Gab (2023), 'About gab.com'. Retrieved from Gab:.
**URL:** *https://gab.com/about*

Google (2023).
**URL:** *https://trends.google.com/trends/*

Hey, A. (2023). Retrieved from GitHub.com:.
**URL:** *https://github.com/alexhey1999/QMUL-MSc-Project.*

Institute, C. (n.d.), 'About'. Retrieved from Cato Institute:.
**URL:** *https://www.cato.org/about*

Jazeera, A. (n.d.), 'Home'. Retrieved from Al Jazeera:.
**URL:** *https://www.aljazeera.com/*

OpenAI (2023), 'Api'. Retrieved from OpenAI:.
**URL:** *https://platform.openai.com/*

Organisation, P. C. (n.d.), 'Uk parties 2019 general election'. Retrieved from Political Compass:.
**URL:** *https://www.politicalcompass.org/uk2019*

Samuel-Azran, T. (n.d.), 'Intercultural communication as a clash of civilizations'. Retrieved from https://www.academia.edu/:.
**URL:** *https://www.academia.edu/30851876/Intercultural_-Communication_as_a_Clash_of_Civilizations*

Stavrianou, A., Brun, C., Silander, T. & Roux, C. (2014), 'Nlp-based feature extraction for automated tweet classification', *Interactions between Data Mining and Natural Language Processing* **145**.

Susser, D., Roessler, B. & Nissenbaum, H. (2019), 'Online manipulation: Hidden influences in a digital world', *Geo. L. Tech. Rev.* **4**, 1.

TensorFlow (n.d.*a*). Retrieved from https://tfhub.dev/:.
**URL:** *https://tfhub.dev/tensorflow/small_bert/bert_en_uncased_L-4_H-512_A-8/2*

TensorFlow (n.d.*b*), 'Classify text with bert'. Retrieved from Tensorflow:.
**URL:** *https://www.tensorflow.org/text/tutorials/classify_text_with_bert*

Unit, E. I. (n.d.), 'Frontline democracy and the battle for ukraine'. Retrieved from Democracy Index 2022:.
**URL:** *https://pages.eiu.com/rs/753-RIQ-438/images/DI-final-version-report.pdf*

Utley, T. (2001), 'I'm v. right-wing, says the bbc, but it's not that simple'. Retrieved from The Telegraph:.
**URL:** *https://www.telegraph.co.uk/comment/4262768/Im-v.-Right-wing-says-the-BBC-but-its-not-that-simple.html*

V.-Dem (n.d.), 'Defiance in the face of autocratization'. Retrieved from V-Dem:.
**URL:** *https://v-dem.net/documents/29/V-dem_democracyreport2023_lowres.pdf*

Xinhua (n.d.), 'Home'. Retrieved from Xinhua Net:.
**URL:** *https://english.news.cn/*

Zhang, L., Ghosh, R., Dekhil, M., Hsu, M. & Liu, B. (2011), 'Combining lexicon-based and learning-based methods for twitter sentiment analysis', *HP Laboratories, Technical Report HPL-2011* **89**, 1–8.

## XIV. Appendices

TABLE XIII
PYTHON LIBRARIES UTILISED

| Library Name | Version Used | Use Case |
|---|---|---|
| alive-progress | 3.1.4 | Training along with other processes can take a long time to run. alive-progress allowed for the use of loading bars |
| beautifulsoup4 | 4.11.1 | When HTML Parsing from the requests module, BS (Beautiful Soup) allows for easy filtering |
| joblib | 1.0.1 | Allows model/file saving for Linear SVM model |
| Nltk | 3.7 | NLTK (Natural Language Tool Kit) allows for multiple pre-processing functions for LinearSVM |
| openai | 0.27.8 | Used for dataset improvement attempt explained in Improvements section |
| pandas | 1.5.2 | Used to create dfs (data frames) for easy data reading |
| Pillow | 8.2.0 | Image Processing Library. Used to load and display the relevant compass image |
| python-dotenv | 0.20.0 | Used to load in .env files Environment variables. Saves pushing sensitive data to GitHub |
| requests | 2.21.0 | Used to connect to web sources |
| scikit-learn | 0.24.2 | Used in LinearSVC learning methods. Used for Classifier and metrics |
| selenium | 4.3.0 | Used in the data collection stage. Uses a Chrome Web Driver to emulate real web driver browsing |
| tensorflow-hub | 0.13.0 | Used for BERT Learning |
| tensorflow-text | 2.10.0 | Used for BERT Learning |
| tf-models-official | 2.10.1 | Used for BERT Learning |

TABLE XIV
BERT X-AXIS MODEL TRAINING STATS FULL

| Epoch | Loss | Accuracy | Precision | Recall | True Positives | True Negatives | False Positives | False Negatives |
|---|---|---|---|---|---|---|---|---|
| **1** | 0.6907 | 0.5647 | 0.6023 | 0.3814 | 1060 | 2078 | 700 | 1719 |
| **2** | 0.4014 | 0.8033 | 0.8969 | 0.6855 | 1905 | 2559 | 219 | 874 |
| **3** | 0.2249 | 0.9057 | 0.9423 | 0.8643 | 2402 | 2631 | 147 | 377 |
| **4** | 0.1671 | 0.9325 | 0.9588 | 0.9039 | 2512 | 2670 | 108 | 267 |
| **5** | 0.1341 | 0.9491 | 0.9744 | 0.9313 | 2588 | 2686 | 92 | 191 |

TABLE XV
BERT Y-AXIS MODEL TRAINING STATS FULL

| Epoch | Loss | Accuracy | Precision | Recall | True Positives | True Negatives | False Positives | False Negatives |
|---|---|---|---|---|---|---|---|---|
| **1** | 0.6747 | 0.6652 | 0.7321 | 0.5208 | 1929 | 3080 | 706 | 1775 |
| **2** | 0.3981 | 0.8136 | 0.9055 | 0.7001 | 1945 | 2576 | 203 | 833 |
| **3** | 0.2080 | 0.9163 | 0.9412 | 0.8880 | 2467 | 2625 | 154 | 311 |
| **4** | 0.1457 | 0.9455 | 0.9558 | 0.9341 | 2595 | 2659 | 120 | 183 |
| **5** | 0.1110 | 0.9581 | 0.9618 | 0.9539 | 2650 | 2674 | 105 | 128 |

TABLE XVI
OPENAI IMPLEMENTATION TWEET CONVERSION

| ID | Original Statement | Tweet Format | Label |
|---|---|---|---|
| **1** | Obama the Born-again Budget Cutter?!? | ”Obama suddenly becoming a budget cutter? Now that’s a plot twist! ???? Let’s hope this newfound austerity leads to some real reforms. #LibertyWins” | Libertarian |
| **2** | North Korea fires two missiles, second test in three days | Another missile test by North Korea... ??? When will we learn the importance of peaceful diplomacy? ?? #EnoughWithTheWarmongering #PeacefulSolutions | Authoritarian |

* Please note that these tweets contain emojis however are not applicable to this report. They have been replaced with question marks.