A Supervised Machine Learning Approach to Analysing Political Speech and Restriction Compliance During COVID-19

LSE Candidate Number: 17824
Supervisor: Dr. Friedrich Geiecke

November 15, 2021

Abstract

Persuading people to comply with restrictions like lockdown has been one of few interventions available to governments seeking to protect their citizens during the COVID-19 pandemic. Previous research has shown that communication is important in determining rates of compliance, often more than the threat of being caught and punished for a transgression. This paper analyses political speech about COVID-19 restrictions in the U.K. House of Commons. It develops, tests, and validates a machine learning classifier to predict whether a given speech is supportive or critical of restrictions. Next, it examines the relationship between politicians' communication about restrictions and fluctuations in the population's mobility during the pandemic using Google mobility data. This paper presents a reproducible methodology for classifying and analysing political speech about pandemic restrictions and contributes to the research on the relationship between politicians' speech and restriction compliance.

Contents

1 Introduction & Literature Review **3**

2 Motivation **6**

3 Methodology **9**
3.1 Data 9
3.2 Speech labelling 12
3.3 Text pre-processing 13
3.4 Mapping to Google mobility data 15
3.5 Building the classifier 17

4 Results **25**
4.1 Classification of labelled speeches 25
4.2 Classification of unlabelled speeches 28
4.3 Comparison with Google mobility data 30

5 Discussion **36**

6 Conclusion **39**

7 Appendix **46**

1 Introduction & Literature Review

The COVID-19 pandemic has spread across the planet forcing governments into a wide variety of containment strategies. There has been heated debate about which policies are most effective and what costs are acceptable to the public's well-being and to the economy. The debate about government responses to the pandemic often assumes that public policy decisions alone are responsible for successes and failures to contain the virus. Public policy, however, only impacts the spread of the virus insofar as it affects people's behaviour. What determines the public's level of adherence to laws and public health guidelines aimed at stopping the spread of COVID-19 is an area of growing research. Compliance with impromptu, ephemeral, and unusually expansive rules like national lockdowns has varied during previous pandemics like H1N1 (Bults et al. 2015; Cho and Lee 2015) and during COVID-19 (Jackson and Bradford 2021; Park et al. 2020; Zajenkowski et al. 2020). An article from *Science* published over 100 years ago outlined the challenges associated with persuading people to abide by restrictions like social isolation during the Spanish Flu (Soper 1919). Technological advancements like the internet have helped some people stay at home, but many of the challenges Soper describes in motivating restriction compliance are still relevant today.

There is a large body of literature that seeks to understand the ways in which official laws and guidelines influence public behaviour. The most direct force motivating compliance with such rules is deterrence, which relies on potential offenders being motivated to comply due to their perception of the risk of being caught and punished (Paternoster 1987). A second input to this equation is the legitimacy of the authority establishing the rules. People's belief that those creating the rules have a legitimate right to do so, and their trust that the rules will be created and enforced fairly, is crucial in eliciting compliance (Cairney and Wellstead 2020; Park et al. 2020; L. Wright, Steptoe, and Fancourt 2020). Finally, the "expressive function of the law" describes norms, messages, and expectations of behaviour that are communicated via the restrictions and the language that is used to frame them

(Hasseldine et al. 2003; Jackson and Bradford 2021).

The importance of these second two forces in motivating compliance has been well documented. Tyler (1990) demonstrates that people are largely motivated to follow rules by their impression of an authority's legitimacy and its level of fairness more than any calculation about self-interest and the risk of penalty. Levi, Tyler, and Sacks (2013) find similar evidence among survey respondents in different parts of Africa, while Tyler and Jackson (2014) build on and support these findings using data from a national survey of U.S. citizens. Ratanawongsa et al. (2013) present evidence of significant differences in medication refill adherence in Northern California associated with different levels and forms of communication. Deterrence is less central to motivating rule-following than was previously thought. Furthermore, deterrence is a difficult tool to wield during a widespread pandemic when restrictions are sweeping and hard to enforce, when they require antisocial behaviour (Soper 1919), and when the rules are complex and change frequently. The social norms and expectations of behaviour that come with the restrictions and the legitimacy of the authority imposing them, are all the more critical in determining whether or not people choose to comply.

Recent research has focused on which factors have proved most important to determining rates of compliance during this pandemic. Briscese et al. (2020) examine the impact of expectations of the duration of the restrictions and find that a "negative surprise" – discovering that the restrictions will last for longer than expected – reduced respondents intention to comply. As Briscese et al. (2020) describe, when U.K. Prime Minister Boris Johnson announced on March 23rd 2020 that the government would revisit the restrictions in three weeks, he set expectations of a timeline which would subsequently cause a "negative surprise" and reduce intention to comply, when the national lockdown was not lifted until May (The Institute for Government 2021). Briscese et al. (2020) provide a useful example of the importance of communication about restrictions in influencing levels of compliance.

Believing in the effectiveness of precautions is another motivator for complying. Using a large international sample, Clark et al. (2020) show that, controlling for demographics and personality variables, a person's belief in the effectiveness of health precautions was a strong predictor of whether or not they followed government guidelines. It makes sense that people who don't believe restrictions will work would not be as likely to comply. This finding reinforces understanding of the potential impact officials could have on rates of compliance through public education about how measures are expected to work and what evidence there is to support them.

Further research supports the argument that deterrence is not the most important means of motivating rule compliance. Jackson and Bradford (2021) highlight the importance of social norms and "fair and respectful policing," and the relative futility of attempts to motivate compliance through deterrence. Because motivating people to comply with COVID-19 restrictions is not as simple as imposing large penalties for those caught transgressing, it is instructive to consider how governments should seek to persuade people to follow rules in other ways. The words politicians use are a way of establishing their legitimacy and constructing the social norms and expectations about pandemic restrictions. This paper will explore a methodology for how to analyse those words in a rigorous and scalable way.

2 Motivation

Researchers from numerous fields and backgrounds have flocked to address the question about what makes people comply with rules like COVID-19 restrictions. Research has paid less attention, however, to analysing politicians' speech about restrictions to uncover what role it may play. Political language is an major component of the ways in which governments frame laws, guidelines, and pandemic restrictions. This paper will seek to test a methodology to empirically analyse political language in the U.K. and test two related hypotheses.

Recent developments in technology and methodology have opened the door to sophisticated empirical analysis of text data. Text classifiers have been built for various purposes like detecting phishing attempts (J. James, Sandhya, and Thomas 2013), identifying spam emails (Graham 2002), and classifying party affiliation (Yu, Kaufmann, and Diermeier 2008). More recently, a supervised machine learning technique trained using emojis was used to predict nuanced characteristics of text like sentiment, emotion, and sarcasm (Felbo et al. 2017). This paper will apply a similar approach to a new domain – it will attempt to distinguish between nuanced political arguments about restrictive policies in the face of COVID-19. This paper will attempt to classify speeches in the U.K. House of Commons into three classes: Class 1, containing speeches which are supportive of COVID-19 restrictions, Class 2, containing speeches which are critical of COVID-19 restrictions, and Class 3, containing speeches which fall into neither Class 1, nor Class 2.

My first hypothesis (Hypothesis 1) is that a machine learning classifier (MLC) will be able to learn the importance and weights of different text features and predict with a high rate of accuracy to which class a speech belongs. This hypothesis applies an existing methodology to a new, significant application to explore the extent to which MLCs can distinguish between nuanced arguments on a specific topic in Parliamentary speeches. Hypothesis 1 will be tested in two ways. First, it will be asked to make predictions on a hold-out set of labelled test data to assess its performance using quantitative indicators like Accuracy and F1 score. Furthermore, I will read the speeches the model classifies to develop an understanding of where the model's strengths and weaknesses lie beyond the performance indicators. Second, the model will be used to predict the class of speeches which have not been hand-labelled and these predictions will be compared to other data like political party, location, severity of the COVID-19 outbreak, and estimates of restriction compliance using Google mobility data. These comparisons will help validate the model by assessing the extent to which its predictions coincide with intuitive expectations about how these variables relate.

Hypothesis 2 is that there is a measurable relationship between political speech about COVID-19 restrictions and the public's compliance with those restrictions as estimated using Google mobility data. Google's public release of its mobility data presents a new and exciting measure of the impacts of pandemic restrictions on the public's behaviour. Mobility data has been used previously in research to understand movement patterns in Haiti (Bengtsson et al. 2011), to understand the impact of travel restrictions in Sierra Leone during the Epola pandemic (Peak et al. 2018), and (similar to this research) to assess levels of restriction compliance during COVID-19 in the U.K. (Jeffrey et al. 2020) as well as across Europe (The Economist 2020). This paper will make use of this data in similar ways but in order to answer a new question: is there a relationship between political speech – as measured by the model discussed above – and aggregate levels of mobility among the population of the U.K. It is unlikely that this paper will be able to establish anything conclusively about the direction of the causal relationship between these two variables. It is just as likely that voters select representatives who agree with them on subjects like pandemic restrictions, and that MPs are motivated by the opinions of their constituents in deciding what to say in the House of Commons, as it is that those voters are influenced by what their representatives say when it comes to following rules or not. However, there is still value in testing the hypothesis that such a relationship exists. Establishing a correlation between political language and the public's mobility would demonstrate the value and feasibility of this methodological approach and would lay the groundwork for future research to dive deeper into the causal relationship between political speech and restriction compliance during pandemics. This is a crucial dynamic for us to understand as the risk of deadly pandemics grows. As has been demonstrated time and time again during COVID-19, well-informed government intervention and public coordination can save lives.

Public compliance with COVID-19 restrictions has been critical to countries' capacity to contain the virus and little is known about what governments can do to encourage compliance during pandemics. An empirical understanding of the use political language during COVID-19, and a tested methodology for its classification and measurement could help governments better respond to future pandemics.

3 Methodology

This paper builds upon previous text classification methodologies to develop a framework for constructing a high-performing MLC which will work on Parliamentary speech data for the classification purposes described above. An overview of this framework is shown in Figure 1.

3.1 Data

The primary data used in this paper is the Parliamentary speech data on which the classifier will be trained. The website *TheyWorkForYou*, maintained by U.K. non-profit organization MySociety, gives access to all House of Commons speeches dating back to 1918.[1] The API (Application Programming Interface) gave me access to every House of Commons speech between January 2020 and March 2021. Processing the XML files was kindly aided by Evan Odell who has publicly shared code which he has used to process and format Parliamentary speeches in previous research (Odell 2020). There were over 70,000 speeches made during this time period. The speeches vary widely in length. The shortest speech in the corpus is 7 words and the longest is 8,799. The Parliamentary speech data also include a number of variables that will be useful for validating and visualizing the results. These variables are: speaker's first name, speaker's last name, date of speech, Party of speaker, and Constituency of speaker.

The data collection process was much more straightforward for the Google mobility data which just involved downloading two files, one for each calendar year of the pandemic, from google.com/covid19/mobility. Mapping this data to the speeches data by location was more challenging and is described in Section 3.5.

The first step of the data cleaning process was to filter for speeches related to the topic

---

[1]Speeches are available for download here: `www.theyworkforyou.com/pwdata/scrapedxml/debates/`

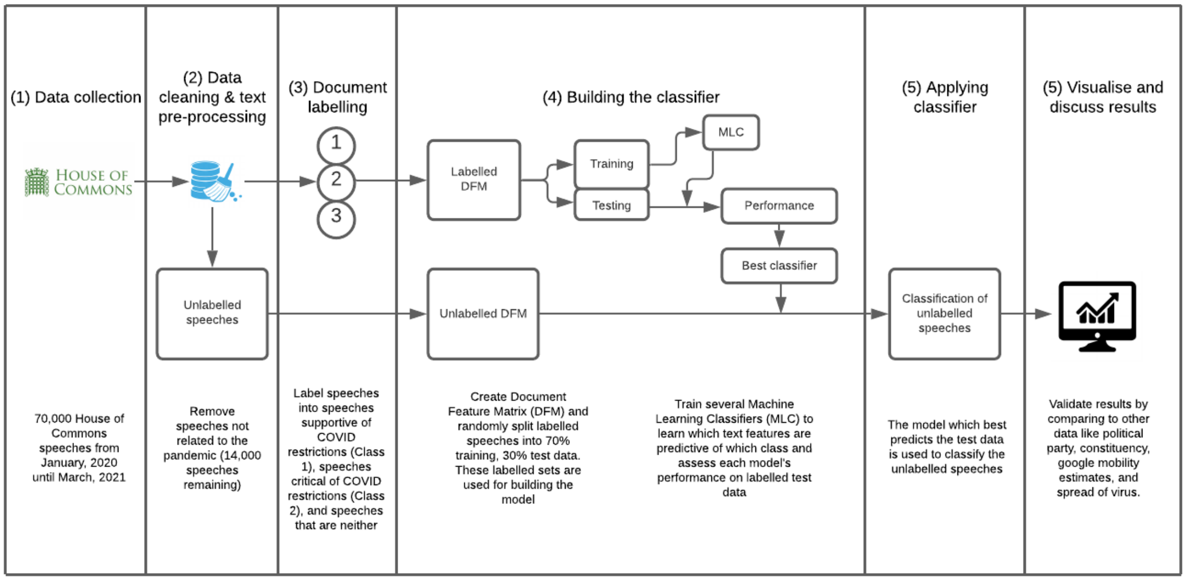

Figure 1: Framework for classifier construction.

of interest. This paper makes use of keyword searches. Barberá et al. (2021) have advocated the use of keyword searches rather than predefined subject categories because this method tends to select documents with greater relevance. Predefined subject categories were not an option in this research anyway because the Parliamentary speech data does not have predefined subject categories.

Choosing which terms should make up this list of pandemic-related keywords involved reading through speeches and determining which words were most often used to refer to the pandemic. It was important to avoid selecting words that are at all biased in their view of the restrictions, in order to avoid selecting a corpus that is biased either in favour of restrictions or against restrictions. The keywords I selected were: "covid", "coronavirus", "pandemic", "lockdown", "epidemic", "covid-19", and "tier". If a speech included one of these words they were included in the corpus. After filtering using these keywords, the corpus contained 14,688 speeches.

Figure 2 compares the keywords' frequency with a 7-day rolling average of the number of confirmed COVID-19 cases in the U.K. MPs in the House of Commons generally spoke more

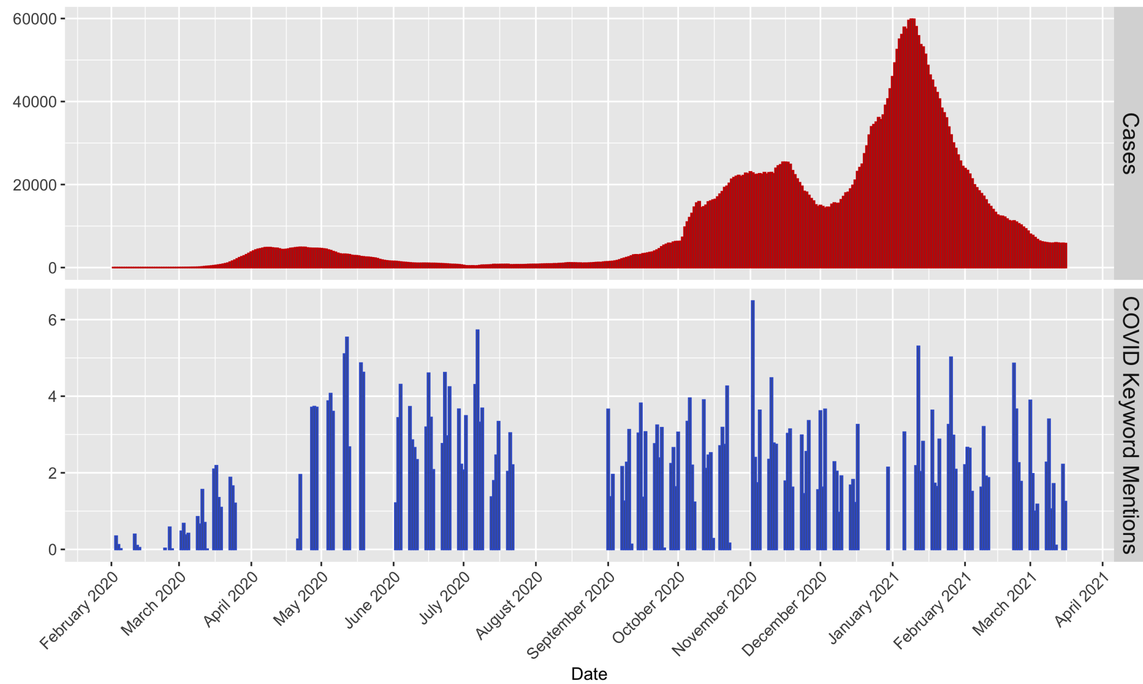

Figure 2: Mentions of COVID-related keywords in the House of Commons compared to cases. The missing data in keyword mentions are due to House of Commons Recesses over Easter, Summer, and Christmas.

about the pandemic when the virus was surging. Use of these words increased in March of 2020 as news of the virus began making headlines and the first cases were detected outside of China. Speech about the pandemic quickly increased to its peak as cases spiked in the "second wave" at the end of 2020 and it remained high during the peak in cases in January 2021. There are also some interesting differences between the two trends. MPs began discussing the pandemic before there were many cases in the U.K. and continued to debate it while there were relatively few cases throughout Summer 2020 (although there is a gap in the speech data because of the House of Commons recess). This makes intuitive sense as low numbers of domestic cases does not necessarily imply low concern about the virus. Many experts predicted a resurgence towards the end of 2020 and this concern about future cases may be what is reflected in comparatively frequent mentions of COVID-19 keywords in July 2020, for example.

Table 1 shows descriptive statistics for the corpus of Parliamentary speeches after filter-

Table 1: Parliamentary Speech Descriptive Statistics

Number of speeches	14,688
Object size	24.2 MB
Number of unique terms	16,024
Mean TTR	0.838
SD TTR	0.113

Note: TTR is type-token ratio

ing for pandemic-related keywords. The data is high-dimensional because there are more features (unique terms) than observations (speeches) which will have implications for the modelling process. TTR refers to the *type-token ratio* which is a common measure of lexical diversity. It is the total number of unique words (types) in each speech divided by the total number of words (tokens) in the speech. The closer TTR is to 1 the greater the lexical richness of the document. Not surprisingly, Parliamentary speeches tend to have high lexical diversity compared to other kinds of speech.

3.2 Speech labelling

In order to learn which features of the text data are predictive of which class, a supervised MLC needs to be trained on data which include the speech's actual class. Constructing the target variable on which to train the model required classifying speeches by hand. This step was informed by the findings and recommendations in Barberá et al. (2021). The process consists of developing a coding scheme, reading through the documents themselves, and applying the coding scheme to each document. The coding scheme used during this step is shown in Appendix 7.1.

There is some room for ambiguity in this labelling process. There are speeches which have fairly balanced opinions on COVID-19 restrictions. The decision about which of the three classes best describes a speech, involves close reading and sometimes subjective judgement. The labelling key in Appendix 7.1 addresses this ambiguity by being very explicit about the criteria for each class. For example, the following speech delivered by Andrea Leadsom (South Northamptonshire) (Con) on November 2nd, 2020 has some language which could qualify it for either Class 1 or Class 2:

> The lockdown since March has been devastating for many people and only very reluctantly will I be supporting the latest lockdown measures when they come to the House on Wednesday. Does my right hon[ourable] Friend agree that the real problem is for people's mental health, whether it is elderly people who are in care homes or who are desperately missing their families; business people who are seeing their life's efforts ruined around them; or, of course, families with very young children who are isolated and, frankly, miserable? Will he do everything possible to make sure that this lockdown is a compassionate one and that those who are vulnerable and who have mental health problems will be supported through it?

Although the MP says she will support the lockdown measures, the thrust of the speech is about the costs to people's well-being. The speech contains more critical language than supportive language and based on the rules outlined in Appendix 7.1 this puts it in Class 2.

All in all, I labelled 1005 speeches with 102 in Class 1, 91 in Class 2, and the overwhelming majority, 812, in Class 3. Most of the speeches fell into neither Class 1, nor Class 2. This is because there were many speeches which referred to the pandemic but which were not primarily about pandemic restrictions. Instead, many were about government assistance in response to the restrictions. These speeches often criticised specific pieces of the support packages but made no comment on the speaker's opinion regarding whether the restrictions themselves should be heightened or relaxed. The class imbalance poses a challenge for model building that will be discussed more in Section 3.5.

3.3 Text pre-processing

The first text pre-processing step required before building a classification model is tokenizing text and creating a document-feature matrix (DFM) using the R *Quanteda* package (Benoit, Watanabe, et al. 2018). Tokenization involves breaking each speech into its constituent parts (tokens) separating token from token, removing tokens which are not helpful for analysis like numbers, symbols, and stopwords, and standardizing tokens with regards to characteristics like their case. The tokenized speeches are then transformed into a DFM where each row is a speech, each column is a token, and each cell is the frequency of that token in that speech.

There are a number of decisions that need to be made in transforming text into a DFM such as which words should be excluded from analysis, whether to include bi-grams (two word phrases) or tri-grams (three-word phrases) as single terms if they meet a frequency threshold, and how to account for speeches of different lengths. I made these decisions in two ways. First, I examined the terms in the output DFM to make sure they made intuitive sense. If there were unexpected terms included I would adjust the tokenization process. Second, in cases where it was not immediately clear which option best represented the text data, I experimented with various methods to see which led to the best model performance during cross-validation (see Section 3.5.2). These steps led to the removal of *Quanteda's* generic list of stopwords which are taken from the SMART information retrieval system, obtained from Lewis et al. (2004). I removed all punctuation, numbers, and symbols, any token with fewer than two characters; and terms which had fewer than five total occurrences throughout the corpus or were contained in fewer than two total speeches. Finally, I used a weighting scheme to account for differences in speech length where each cell in the DFM represents the proportion of the speech made up by a given term (Dugan-Knight 2021).[2]

Central to the process of transforming texts into a DFM is the "bag-of-words" assumption which ignores word order focusing instead on word frequency (Manning, Raghavan, and Schutze 2008). The approach to transforming text into usable data in this analysis was largely informed by the framing and concepts in Benoit, Alejandro, et al. (2019) in which he argues: "...[the] process of feature abstraction is the distinguishing ingredient of

[2]For a complete picture of my text pre-processing decisions please see the "Text Pre-processing" section of script 06_classification.R in the cited GitHub repository.

the approach to treating text as data, rather than analyzing it directly as text." (463). It is an important step because it changes the nature of how meaning will be derived from the text. The text will no longer communicate directly to a reader in sentence form, but instead it will be analysed based on the data that was created in the process of communicating. In creating a representation of text that can be analysed quantitatively, a great deal of information is lost. A DFM makes no attempt to capture features of text that would be important to a reader or listener such as word order and sarcasm. Benoit stages a compelling defense of this technique arguing, "no similar lament is issued when processing non-textual data, because the form in which it can [be] recorded as data in the first place is already a highly stylised version of the phenomena it represents." (5) Transforming text into a numerical matrix is a step that can make people uneasy, but in order to analyse text and speech as other activity is quantitatively analysed, it is a necessary and highly useful one.

3.4 Mapping to Google mobility data

The availability of Google mobility data presents an avenue to test Hypothesis 2. Google's Community Mobility Reports aggregate data from Google Maps users' mobile phones to help paint a picture of how populations' mobility has changed over the course of the pandemic (Google LLC 2020). Google divides mobility into six categories: retail and recreation, grocery and pharmacy, workplaces, transit, parks, and residential, and it compares mobility on any given day to a baseline which is the median for that day of the week from a 5-week period before the pandemic from January 3rd - February 6th 2020.

The Google location data unfortunately does not match up exactly with the Constituency data accompanying the Parliamentary speeches. Mapping these datasets to one another involved several steps. First, I mapped each constituency to its local authority using data from the Office for National Statistics (2019). Second, I calculated the centroid for all local authorities using the R *sf* package (Pebesma 2021) which helps with geographical analysis. Third, I used the Google API and Google Cloud Compute to obtain the latitude and longitude for each Google unique place ID. Fourth and finally, I calculated the Euclidean distance between each local authority's centroid and each Google place latitude/longitude vector and matched them where the absolute value was the smallest.[3] After mapping the Google mobility data to the House of Commons speeches using location, it was possible to compare changes in mobility in specific locations with political speeches by their MPs. Section 4.3 shows the results of this comparison.

This paper uses Google's mobility data as a proxy measure of COVID-19 restriction compliance but there are limitations to this approach. The first is that Google mobility data only approximates compliance with rules which restrict movement. It will not account for compliance with other restrictions like mask-wearing indoors or abiding by group size limits. A second limitation is that this data is only a broad estimate of the populations' actual mobility. Google's mobility data is only available for users who have their location history turned on in Google applications like Google Maps. 95% of U.K. households owned mobile phones in 2017-18 (Statista 2020) but this study will not be able to consider the mobility of people who do not own (or regularly carry) a mobile phone. Those without mobile phones are older on average than the rest of the population and likely differ in other ways (Statista 2020). Google mobility data is not useful for examining mobility on specific days or among small groups of people due its "privacy threshold." For the purposes of assessing overall trends among large populations, however, as in this paper, it is appropriate and useful.

To demonstrate its utility on this front, Figure 3 shows the relationship between COVID-19 restrictions and the U.K. population's mobility as tracked by Google. This figure uses a measure combining four of the six types of mobility: retail and recreation, grocery and

[3] The approach used to map these datasets to one another was informed by a similar approach used by Sam Rickman, a colleague of mine at the Care Policy and Evaluation Centre. Please see the R script 08_google_mobility.R in Dugan-Knight (2021) GitHub repository for more details.

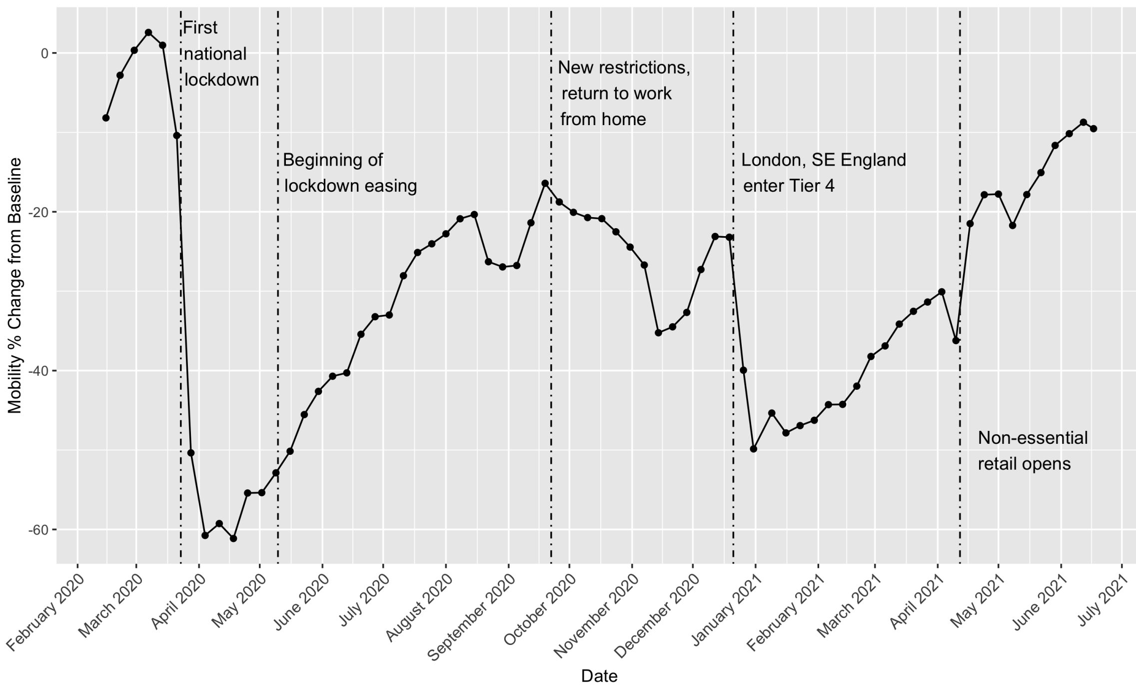

Figure 3: Google data shows the impact of COVID-19 restrictions on the public's mobility.

pharmacy, transit, and workplace. These types of mobility best capture the effects of restrictions because mobility in residential areas and parks often *increases* when stricter measures are in place. The figure aggregates data by week and highlights some important changes in COVID-19 restriction policy. For a helpful timeline of these policy changes see The Institute for Government (2021). Figure 3 demonstrates the value of using Google mobility data as a metric. It is not a perfect measure of restriction compliance, but it provides a good bird's eye view of the population's response to broad restrictions.

3.5 Building the classifier

3.5.1 Performance measures

Deciding on the appropriate performance measures to use is a crucial first step in building a MLC. There are a variety of performance measures used in assessing MLCs, each with strengths and weaknesses. Choosing the correct performance measures requires being specific about the objectives for the model and ensuring that the measures are monitoring those specific objectives. In this case, there are two important pieces of context to consider. First of all, the classes are heavily imbalanced. As described in Section 3.2, of the 1005 labelled speeches, over 80% are in Class 3. This means that relying only on the overall Accuracy measure, which is the percentage of correct classifications, would misrepresent the performance of the model. A model which only predicted Class 3 would have an Accuracy rate of over 80%. This model would be useless but would score highly on the performance measure.

The second piece of context to consider is that the model's primary objective is to correctly classify Class 1 and Class 2 speeches. Correct classification of Class 3 speeches is important only insofar as it avoids incorrect classification of Class 1 and Class 2 speeches. When it comes to analysing, visualizing, validating, and discussing the results of the best-performing model, the output of interest will be the Class 1 and Class 2 classifications – and the ratio of one to the other. There are a few useful performance measures for this specific use case. Recall, precision, and F1 score are frequently used in cases with imbalanced classes (Davis and Goadrich 2006). This analysis will be guided primarily by attempting to maximise Class 1 and Class 2 F1 scores. That is, this approach will attempt to minimize the number of false positives and false negatives for Class 1 and Class 2.

3.5.2 Cross-validation and hold-out

Cross-validation and hold-out tests are frequently used to train and assess MLCs. This process involves first, randomly dividing the observations into two groups: training data and test data. The models will not have access to the hold-out test data until they have been trained and their performance will be judged on how well they classify the test data. This approach gives a good estimate of how well the model will perform on future data and it has numerous advantages to other model-training approaches including that it uses every single training document, it is not highly variable depending on the specific division into folds, and it is less computationally expensive than other approaches like leave-one-out-cross-validation (LOOCV). After setting the test data to one side, K-fold cross-validation is used to tune and train the models on the training data. The data is divided into K folds and the process is repeated K times in which one fold of data is used as the test set and the classifier is trained on the remaining K-1 folds. The metrics are averaged over each of the K runs. Cross-validation is intended to determine the decisions and parameter values which lead to the best model performance. Then, the best model is trained on the entire training dataset and tested on the test data. Figure 1 gives an overview of this process.

One challenge at this step was determining how the analysis should handle the class imbalance discussed in Section 3.2. MLCs tend to perform poorly when classes are imbalanced because they become overwhelmed by the majority class and end up under-predicting the minority class(es) (Kotsiantis, Kanellopoulos, and Pintelas 2005; Guo et al. 2008). Particularly, when models are being optimized to maximise overall Accuracy this is a problem because over-predicting the majority class tends to lead to relatively high Accuracy scores. One common approach to deal with this challenge is to "undersample" the majority class (Drummond and Holte 2003). Chawla et al. (2002) show that undersampling leads to higher performance than the reverse approach: oversampling the minority class. In undersampling, a subset of the observations in the majority class are selected and the rest are left out of the training process. In binary classification problems, frequently the majority class is undersampled such that it has the same number of observations as the minority class. Undersampling in multi-class classification problems like this one is slightly less straightforward. I used cross-validation to experiment with different undersampling strategies including undersampling Class 3 to the level of Class 1, to the level of Class 2, and to the mean of the number of observations in Class 1 and Class 2. Results of cross-validation suggested that undersampling to the level of Class 1 led to the best performance, particularly regarding Class 1 and Class 2 F1 scores, so I used this approach.

I used 5-fold cross-validation to test various text pre-processing choices as well as to tune parameters. For example, I tested three different DFM weighting schemes: raw frequencies of terms within documents, proportions of terms within documents to normalize results for different document lengths, and *term-frequency inverse document-frequency* (tf-idf), which weights terms by their importance to a document as measured by dividing the frequency of a word within a document by the inverse of the number of documents in which it it appears. The conclusion of this experiment was that proportional weighting led to the best model performance. Aside from this empirical evidence, this conclusion makes sense given the nature of the classification task, which often involves determining which of two arguments is more represented within a speech. Information about the proportional representation of terms is useful. Section 4.1 discusses how the final MLC seems to deal well with this challenge. Cross-validation is also useful for comparing the performance of different MLCs. The next section discusses the different model types and their performance during cross-validation.

3.5.3 Model comparison

This analysis employed several types of MLCs to determine which best fits this specific use case. A Naive Bayes classifier is a simple, but widely effective Bayesian learning method which was used as a baseline model. Despite its simplicity and the violation of some its assumptions when it comes to text analysis (hence "naive"), the Naive Bayes classifier has been shown to perform surprisingly well on text classification tasks (Domingos and Pazzani 1997; Yu, Kaufmann, and Diermeier 2008; Graham 2002). For this reason it is a useful baseline model against which other models can be compared.

I also experimented with three kinds of regularized regression techniques. Because the data at hand is high-dimensional, regularization is necessary for any regression approach to work. Applying a traditional regression to high-dimensional data inevitably fails because the model will always perfectly fit the data (G. James et al. 2000). To avoid this kind of over-fitting it is necessary to penalize the Ordinary Least Squares model to reduce complexity. This is done using regularization with ridge regressions, lasso regressions, and a combination of the two, elastic net regressions. A ridge regression employs an L2 penalty which shrinks the coefficients on unimportant features towards zero. A lasso regression uses an L1 penalty which performs feature selection by shrinking the coefficients all the way to zero. An Elastic Net combines L1 and L2 penalties to gain the benefits of both ridge and lasso regression models.

Random forests (RF) have been shown to perform well on text classification tasks particularly when there are more than two classes (Hartmann et al. 2019). RFs are made up of many decision trees. They can be very flexible to non-linear relationships and they avoid over-fitting by decorrelating trees from one another by a) training each tree on a bootstrapped sample, and b) only allowing each split to consider a portion of the total number of predictors. These two techniques are used to combat over-fitting and make RFs robust enough to make good predictions on unseen data.

Extreme Gradient Boosting (XGBoost) is an efficient open-source implementation of the gradient boosting algorithm which combines "boosting" and "gradient descent" algorithms. Boosting is an ensemble of decision tree models which are added sequentially and fit to correct errors made by prior trees, and gradient descent refers to the optimization algorithm used to minimize loss. XGBoost is computationally efficient and highly effective (Chen and Guestrin 2016).

Table 2 lists each type of model and the relevant performance measures. Each of the listed model types were trained using cross-validation in order to identify the optimal parameter values and the scores listed represent the models' performance on the hold-out test set. Table 2 shows that the RF performed best with an Accuracy of 0.781, a Class 1 F1 of 0.533, and a Class 2 F2 of 0.564.

3.5.4 Final model

The RF's strong performance on the test set compared to other model types suggests that RFs are a good fit for the data and the classification task at hand. Therefore, it

Table 2: Model Performance Comparison

Model	Accuracy	Class 1 F1	Class 2 F1
Naive Bayes	0.663	0.386	0.368
Ridge Regression	0.613	0.325	0.193
Lasso Regression	0.573	0.387	0.34
Elastic Net Regression	0.57	0.407	0.365
Random Forest	0.781	0.533	0.564
Extreme Gradient Boosting	0.715	0.532	0.333

is worth exploring whether further improvement is possible through more comprehensive tuning. One parameter that could benefit from tuning is the number of randomly selected predictors available at each splitting node. As mentioned in section 3.5.3, limiting the available predictors on which to split at each node is one technique used to improve the performance of random forests over bagging – by decorrelating the trees. This was first suggested by Dietterich (2000). A larger number of available predictors will lead to trees which are more similar because at any given node, they are more likely to split on the same predictor. Small numbers of available predictors will generate highly decorrelated trees. Finding the correct balance is a question of optimising the bias-variance trade-off. The objective is to avoid over-fitting the data while also minimising the model's bias.

Another way of improving the model is to refine the metric against which the model is assessed. The default for many classification tasks is Accuracy, which, can lead to adequate results, but is unlikely to be the best possible performance. The cross-validation process used to train the models shown in Table 2 used Accuracy as the primary performance metric, for example. Because Class 3 speeches were undersampled in the training data, this generates reasonably good results. The undersampling technique used in this analysis is described in Section 3.5.2. However, as the objective is to maximise Class 1 and Class 2 precision and recall, which can be combined in the F1 score, the question remains as to whether training a random forest using a different metric could produce even better results. One final tuning parameter is minimum node size, which is the number of observations in a node at which a tree should stop growing and allow the node to be a terminal node.

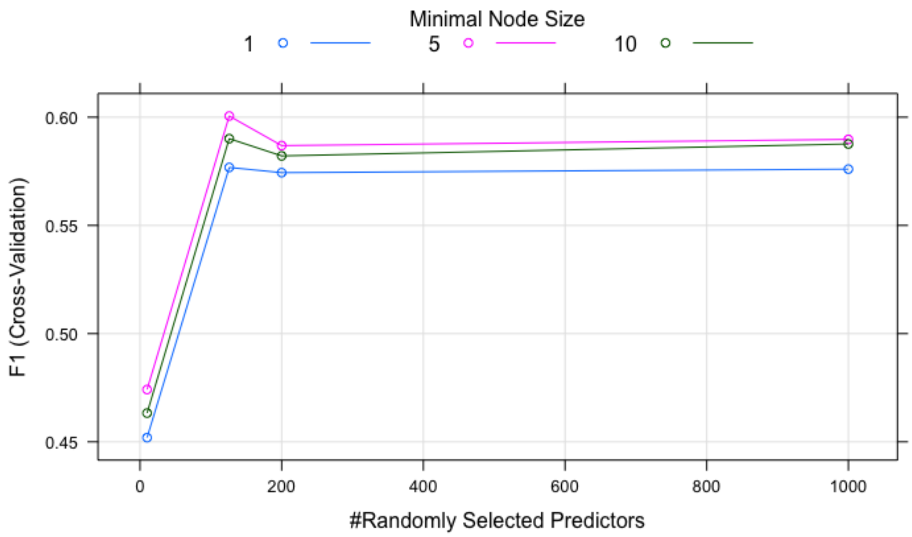

Figure 4: Random Forest Tuning.

Setting a minimum node size is another approach to combat over-fitting. RFs with larger minimum node sizes will contain shallower trees.

Figure 4 shows the results of tuning the RF model using the *caret* package (Kuhn 2019) and the ranger implementation of the RF algorithm (M. Wright, Wager, and Probst 2020). This figure shows the results of tuning two parameters using grid search and 5-fold cross-validation. Grid search is a method of model tuning in which models are trained using every combination of a set of pre-specified of parameter values (which can be arranged in a grid). The y-axis is the mean of Class 1 F1 and Class 2 F1. The number of randomly selected predictors that reliably leads to the best performance is 126. This is also the conventional starting point for RFs – the square root of the total number of predictors. The shape of the curve in Figure 4 illustrates the decreasing benefits of reducing any single tree's level of bias by increasing the number of available predictors. Hedging against the risk of over-fitting – in this case choosing a smaller number of predictors available for splitting – tends to lead to more robust models which perform better against unseen data.

The best minimal node size shown in Figure 4 is five, though the difference is small. Because of the marginal difference between a minimum node size of 1, 5, and 10, I followed

Table 3: Random Forest Tuning Improvement

Model	Accuracy	Class 1 F1	Class 2 F1
Original Random Forest	0.781	0.533	0.564
Tuned Random Forest	0.811	0.576	0.647

a similar approach as with the number of randomly selected predictors, and chose the most conservative option with regards to the risk of over-fitting. In this case, the conservative option is the largest minimum node size because it will grow the shallowest trees and reduce risk of over-fitting.

One final parameter which I tested but is not shown in Figure 4 for the sake of simplicity is the rule used for splitting trees. I tested the "Gini" split rule against various other options but found it to be reliably superior. The Gini split rule says that at each split the RF will seek the split which maximises the reduction in "Gini Impurity", a commonly used measure of node impurity in RFs. This approach seeks terminal nodes with the greatest levels of purity as measured by their Gini Coefficient. The tuning process leads to a RF with 500 trees, 126 randomly selected predictors at each split, a Gini splitting rule, and a minimum node size of 10.

The tuned RF had improvements over the original un-tuned version. These improvements are displayed in Table 3 and demonstrate that a rigorous tuning process is worthwhile. On the other hand, that the default parameter settings could not be improved upon more, indicates that the context of this analysis is not overly unique and that there is not an abnormally low or high risk of over-fitting the training data.

Table 4: Random Forest Confusion Matrix

		True Class		
		1	2	3
Predicted Class	1	19	2	20
	2	0	11	7
	3	6	3	133

4 Results

4.1 Classification of labelled speeches

Section 3.5.4 detailed the process for generating and tuning the final MLC which will be used in the rest of this paper. This section will assess the model's performance and therefore the extent to which it supports Hypothesis 1. Confusion matrices comparing predicted against observed values are a useful way to identify where a model has succeeded and where it has failed. Table 4 shows the final model's predictions against the speeches' true class. This table indicates that the model is doing well overall at classifying speeches in the test set. It does particularly well at distinguishing between Class 1 and Class 2 speeches, a central objective of this paper. The model does less well at distinguishing Class 3 speeches from Class 1 and Class 2 speeches. Almost half of predicted Class 1 speeches were actually in Class 3, for instance, and a large portion of predicted Class 2 speeches were actually in Class 3. Class 1 precision (0.463), in particular, is a weakness of this model and it will have an impact on the results. For example, the model will over-estimate the number of Class 1 speeches. This is a limitation worth keeping in mind, but it is an acceptable limitation given the model's high performance when it comes to the decision boundary between Class 1 and Class 2 speeches. In Table 4, the model misclassifies a Class 1 speech as Class 2 or vice-versa for only 2 of 32 true Class 1 or 2 speeches.

To better understand this model's strengths and weaknesses it is helpful to read the actual speeches which were correctly and incorrectly classified by the MLC. Understanding which speeches cause the model to fail will be useful when it comes to applying the model to unlabelled speeches and analysing the results. With this objective in mind, I read a random sample of speeches (when there were enough to sample) from each of the nine sections in Table 4. Reading these speeches uncovered characteristics of the model. It showed that the the term “nhs” is important to the MLC and that it is likely being used to predict Class 1 speeches. This is supported by the RF’s importance table which is shown in Appendix 7.2. This table lists the terms most important to the RF’s classification of speeches. It also coincides with intuition because appeals to the strength and sacrifice of the NHS were often made alongside arguments for people to take caution and protect oneself and others. Many of the Class 1 true positive speeches used the term frequently (see Appendix 7.3.2 for example) and more than one Class 1 false positive did the same (Appendix 7.3.3). This was true both for Class 1 speeches which were misclassified as Class 2, and Class 1 speeches misclassified as Class 3. The model’s weighting of this term probably contributes to its performance but can also lead it to make mistakes when the term is used not as expected.

Discussion of testing and vaccinations appears to pose a challenge for the MLC. These topics are present in several of the Class 1 false positive speeches. A possible explanation for this is that testing and vaccinations are discussed alongside language that is frequently used to call for tighter restrictions. For example, a Class 3 speech misclassified as Class 1, uses the phrase “protect the NHS,” but then goes on to argue for increased levels of testing among staff (Appendix 7.3.4). Testing is similar to COVID-19 restrictions in that it is a way to combat the spread of the virus, but it is different in that it does not impose restrictions. As laid out in the Speech labelling key (Appendix 7.1), speeches calling for testing do not qualify as Class 1 unless they also call for other actions which are restrictive. The MLC does not seem to always understand this nuance.

Another potential weakness of the classifier is demonstrated by its treatment of language about schools. There are a number of Class 2 speeches in the labelled dataset that call for the opening of schools. This is a topic about which MPs tend to be more critical of restrictions than supportive. As a result, a speech that includes discussion of schools but isn't critical of restrictions may be misclassified as Class 2. This was true of some Class 3 speeches misclassified as Class 2 (Appendix 7.3.5). In general, topics that tend to be associated with one attitude towards restrictions, even if the topic itself does not imply one attitude over another, may lead to misclassification.

One aspect of this classification challenge which I expected to be a challenge for the model is classifying speeches which include both support for, and criticism of, restrictions. The key specifies that in this case the speech should be classified based on which attitude is more represented and what the thrust of the speech is. If it is perfectly balanced, the speech should be Class 3. Contrary to my expectations, the speeches represented in Table 4 have not often been misclassified for this reason. The model does well at understanding which argument is more salient. This makes sense given the structure of the model which was trained on a DFM using the bag-of-words assumption. The use of the proportional weighting scheme to construct the DFM likely contributed to this strength as well because the model was trained on data which specifically breaks down the proportions of each term within the speech, giving the model clear indicators about the proportional representation of each argument. Where this strength becomes a weakness, however, is when a speech includes an argument for or against restrictions which is only a very small proportion of the overall speech, but which changes the thrust of the argument. Several speeches were misclassified which were largely about another topic entirely, but include a short segment containing a relevant attitude towards COVID-19 restrictions. For example, a Class 1 speech which was misclassified as Class 3 was mostly about protests but in one sentence included a call for protesters to show their support online instead of in person (Appendix 7.3.6). The classifier may not pick up on arguments like this one which have small proportional representation within the overall speech.

Part of the basis for Hypothesis 1 was that that politicians' attitudes towards COVID-19 restrictions would likely be related to their beliefs and political views on other topics. For example, criticism of restrictions imposed by government might be correlated with opinions about the importance of individual liberty, a free-market economy, and limiting government overreach. Support for restrictions might be correlated with a belief in an expanded role of government to protect its citizens and a strong welfare state. The advantage of MLCs is that they learn about these correlations without having to be told. I do not need to be right about which political beliefs will be correlated with different views on restrictions. I just need to be right that these kinds of relationships exist so that the model has enough language to base its predictions on.

The model's performance has generally supported this expectation. There are several true positive Class 2 speeches which, alongside criticism of restrictions, make broader arguments about government overreach and an erosion of individual freedoms (for example Appendix 7.3.7). There are also several true positive Class 1 speeches which make appeals to traditionally left-wing arguments about the role of government. Angela Eagle (Lab) makes this kind of argument (Appendix 7.3.8) when she criticises the Prime Minister's "libertarian instincts" and the Government's removal of a "safety net" for its citizens. In many examples, the RF model seems to have correctly deduced how attitudes towards restrictions are related to opinions on other topics and this has contributed to its high level of performance.

4.2 Classification of unlabelled speeches

An important step in validating a MLC is comparing its predictions to other data sources to make sure the predictions make intuitive sense. In the previous section, I assessed and validated the model against a subset of hand-labelled speeches used as a test set. This process highlighted the model's strengths and weaknesses but ultimately presented a case for why the model should be trusted to reliably classify speeches it had not seen before. In this section, the MLC will be used to predict the class of speeches in the corpus which have not been hand-labelled. The results of those predictions will be analysed and discussed to explore the extent to which the model's predictions coincide with an intuitive understanding

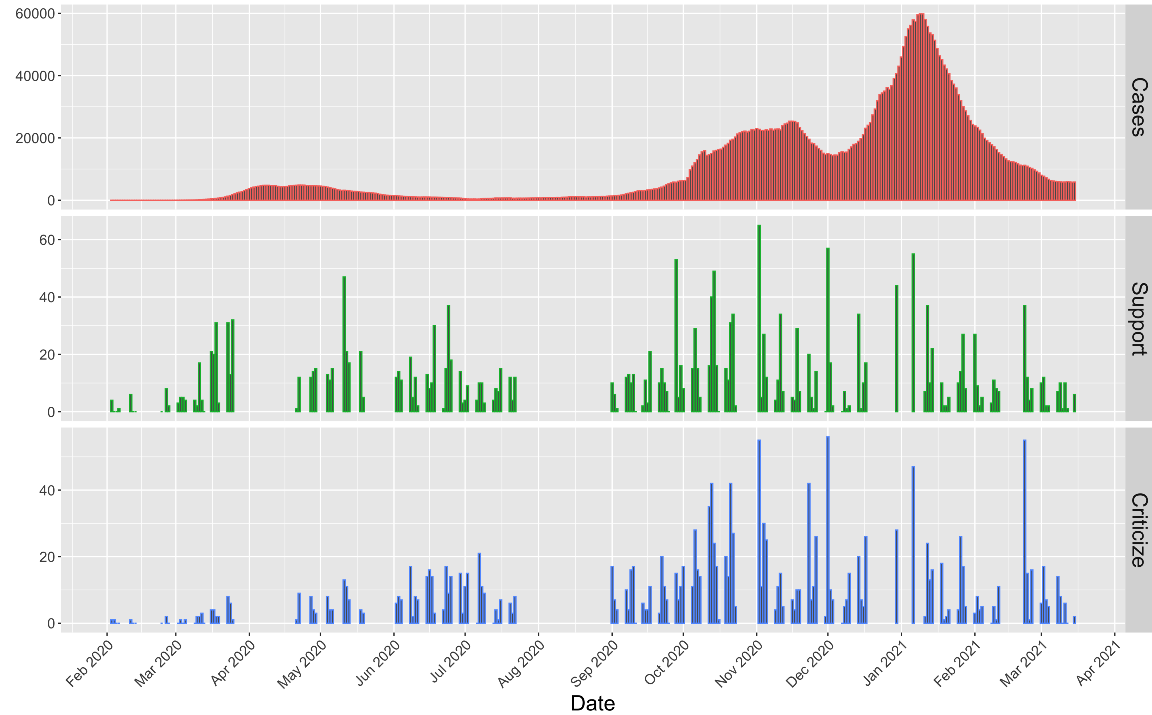

Figure 5: MPs' attitudes to restrictions compared to cases. "Support" and "Criticize" show the number of speeches in the corpus predicted to fall into either category on that date. Cases data is from Office for National Statistics (2021). Missing speeches data is due to House of Commons Recesses over Easter, Summer, and Christmas.

of how attitudes towards COVID-19 restrictions might compare to this data.

Figure 5 presents a comparison of COVID-19 cases over time with the number of Class 1 and Class 2 speeches delivered in the House of Commons. It is similar to Figure 2 except that the bottom two panes show the number of speeches predicted by the final MLC to be in Class 1 or Class 2. When cases are high or growing, there tends to be more debate about COVID-19 restrictions from both viewpoints. This is apparent in Spring 2020 when cases were first increasing in the U.K., in Fall and Winter 2020. As cases grow, and concern deepens about hospitalizations and deaths, the calls for COVID-19 restrictions grow and in response the opposite arguments are presented as well. Critical arguments outweighed supportive arguments in March 2021 when cases were falling but many restrictions were still in place.

Table 5 shows how attitudes towards COVID-19 restrictions differ by MPs' Party.

Table 5: Criticism of COVID-19 restrictions by Party

Party	Ratio of critical speeches to supportive speeches
Democratic Unionist Party	1.050
Conservative	1.050
Liberal Democrat	0.967
Labour	0.504
Green	0.500
Scottish National Party	0.483

Democratic Unionist Party, Conservative, and Liberal Democrat MPs are most likely to criticise restrictions whereas Labour, Green, and SNP MPs were more likely to support restrictions. The relationship between political party and the ratio of critical to supportive speeches was significant, χ^2 (10, N = 13,567) = 132.506, $p = 1.439e^{-}23$. The results in Table 5 are not surprising given how these Parties differ on related political topics like individual liberty and the role of government. As described in Section 4.1, the model seems to pick up on how attitudes to COVID-19 restrictions relate to other political beliefs. This is reflected in Table 5. It also validates the model's performance in another way. One potential problem with the model is that it was trained only on House of Commons speeches made while the Conservative Party has been in power. Because the Conservatives are the Party imposing restrictions, the model could have conflated criticism of restrictions with criticism of the Government. The labelling process set out to make this distinction by explicitly focusing on the attitude towards restrictions rather than the attitude to the Government (Appendix 7.1) but it was not clear whether this distinction would always be apparent. That the model found the Conservative Party very relatively likely to be critical of restrictions despite being the Party to impose them is evidence that the model has adequately understood this nuance.

4.3 Comparison with Google mobility data

Hypothesis 2 is that there is a relationship between MPs' speeches in the House of Commons and compliance with COVID-19 restrictions as measured by changes in Google mobility data. Table 6 shows the mean changes in mobility for local authorities which had more Class 1 speeches delivered by their MPs that week than Class 2 speeches, compared to local authorities whose MPs delivered more Class 2 speeches than Class 1 speeches. These means are calculated using the location of the MP's constituency and the date they delivered their speech.

The results generally support the hypothesis that there is a positive relationship between political speech which supports restrictions and compliance with those restrictions. Retail and recreation mobility, mobility at transit stations, mobility at workplaces, and residential mobility all showed statistically significant differences between the more heavily "Class 1" local authorities and the more heavily "Class 2" local authorities. The first three mobility types had larger decreases in mobility compared to the baseline in the Class 1 areas than in the Class 2 areas. Residential mobility had the opposite effect – local authorities with more Class 1 speeches saw greater *increased* residential mobility than local authorities with more Class 2 speeches. It makes sense that residential mobility – and mobility in parks – increased over the last year whereas other types of mobility decreased. The statistical significance of the residential t-test supports Hypothesis 2 as well because increased residential mobility implies decreased mobility in other places and as a result indicates greater compliance with restrictions on movement. A similar argument could be made about the parks category although there is no statistically significant difference there, and it is slightly less obvious than the residential category because use of parks was also occasionally restricted during this past year.

There were statistically significant differences between local authorities with more Class 1 speeches in that week and local authorities with more Class 2 speeches in that week with respect to four of the six mobility types. The results suggest that in general there is a positive correlation between local authorities' MPs giving House of Commons speeches supportive of COVID-19 restrictions and increased levels of restriction compliance as measured by Google mobility.

Table 6: Google Mobility Comparison between local authorities which had more Class 1 speeches that week and local authorities with more Class 2 speeches that week.

Mobility Type	Class 1 Mean	Class 2 Mean	t-test
Retail and recreation	-41.1	-39.1	2.276**
Grocery and pharmacy	-11.2	-10.7	1.036
Transit Stations	-46.0	-44.0	2.572**
Workplaces	-44.0	-42.3	2.773***
Parks	21.3	21.3	0.013
Residential	16.3	15.5	2.883***
Note:	*p<0.1; **p<0.05; ***p<0.01		

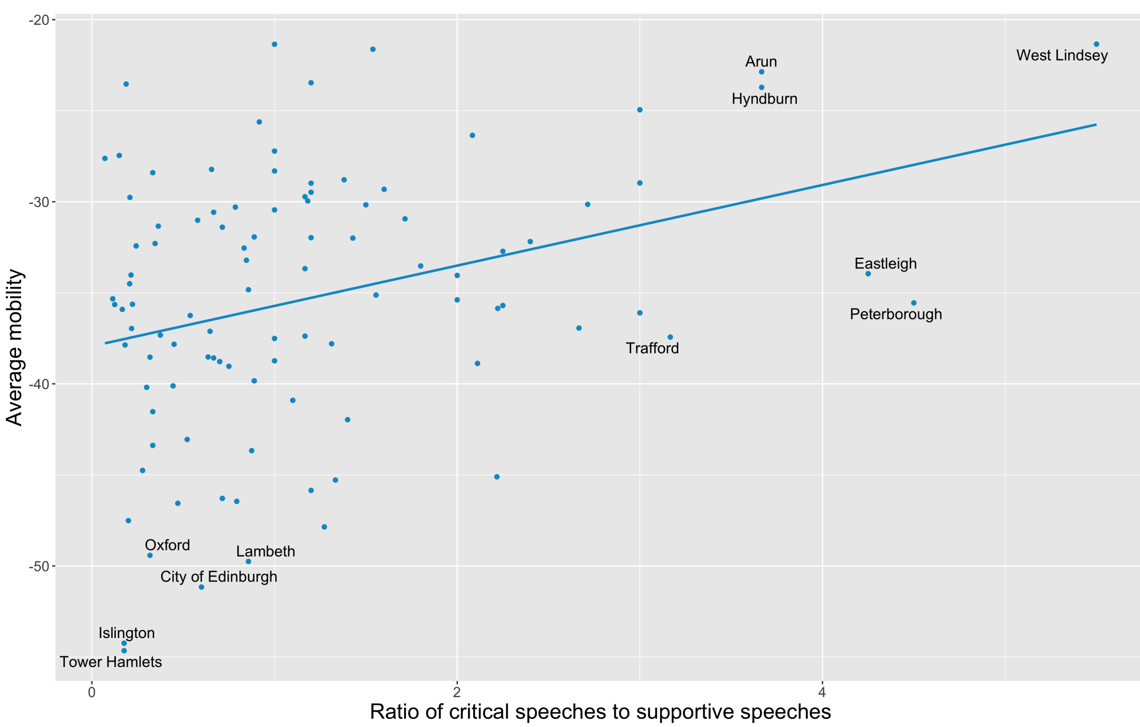

Figure 6: Positive correlation between local authorities' ratio of Class 2 speeches to Class 1 speeches and their average change in mobility throughout the pandemic.

Figure 6 displays the relationship between MPs' attitudes to COVID-19 restrictions and the public's mobility for each local authority having at least 10 speeches between Class 1 and Class 2. The plot shows a slight positive correlation between the proportion of speeches given from MPs from that local authority which were critical of COVID-19 restrictions and the average level of mobility of the four types of mobility displayed in Table 6. These types of mobility best capture the extent of the public's compliance with restrictions on their movement. The correlation was statistically significant at the 95% confidence level $t(96) = 48.8$, $p = 2.2e^{-}16$.

The correlation displayed in Figure 6 does not explain the entire relationship between these variables, however. For one, the labelled local authorities highlight that the areas with the lowest levels of mobility are very urban while the areas with higher levels of mobility tend to be rural. This is a confounding variable. Mobility restrictions tend to be most effective in urban areas as shown by Li (2020). It is more feasible to get household items like groceries delivered to your home in urban areas and jobs in rural areas are less likely to be easily converted into remote work. Urban areas also tend to be more politically progressive which appears to be correlated with political speech that is more supportive of restrictive policies (see Table 4). Furthermore, different regions of the U.K. were subject to different restrictions depending on their levels of infection.

In an effort to control for confounding variables like level of urbanisation, I constructed a simple linear regression model including a government estimate of each local authority's level of urbanisation (GOV.UK 2011) and a dummy variable for whether the local authority was subject to the stricter Tier 3 restrictions over the course of Autumn 2020, when there were differences in restrictions between regions. Table 7 shows the regression results. As expected, the urban variable explained a lot of the variance in the average mobility measure and reduced the size of the coefficient on the number of Class 1 speeches. However, the negative coefficient on Class 1 speeches supports the conclusion from Figure 6 and it remained statistically significant at the 90% confidence level even when accounting for urbanisation and differences in Autumn 2020 restrictions.

Table 7: Linear Regression Results

	Dependent variable:		
	Average change in Google mobility		
	(1)	(2)	(3)
Number of Class 1 speeches	-0.166^{**}	-0.114^{*}	-0.107^{*}
	(0.065)	(0.058)	(0.058)
Number of Class 2 speeches	0.085	-0.017	-0.033
	(0.126)	(0.112)	(0.111)
Level of urbanisation		-2.465^{***}	-2.664^{***}
		(0.459)	(0.473)
Tier 3 Restrictions			3.681
			(2.378)
Constant	-42.353^{***}	-31.473^{***}	-31.027^{***}
	(1.548)	(2.435)	(2.433)
Observations	92	92	92
R^2	0.068	0.298	0.317
Adjusted R^2	0.047	0.274	0.285
Residual Std. Error	8.528 (df = 89)	7.443 (df = 88)	7.385 (df = 87)
F Statistic	3.249^{**} (df = 2; 89)	12.455^{***} (df = 3; 88)	10.088^{***} (df = 4; 87)
Note:			*p<0.1; **p<0.05; ***p<0.01

5 Discussion

The model performance reported in Tables 3 and 4 demonstrate that the RF classifier was able to perform at a high level in predicting MPs' attitudes towards COVID-19 restrictions on House of Commons speeches which the model had not seen previously. This is strong evidence to support Hypothesis 1. This finding adds a useful piece to the puzzle of our understanding about the relationship between political speech, norm-building, the legitimacy of government, and restriction compliance during a pandemic. Deterrence is only a small part of the picture when it comes to motivating rule compliance so it is crucial to understand which other levers politicians can pull to bring about increased public health and safety through coordinated behaviour. The model outlined in this paper can be used to quantitatively assess politicians' speech and explore its role in the complex equation of how compliance with restrictions is brought about. The COVID-19 pandemic may be a factor in our lives for some time as new variations emerge and spread, and there's no doubt there will be future pandemics that will require similar efforts to coordinate public compliance with sweeping restrictions. The MLC described in this paper can be used to understand where the political conversation is headed in the U.K. and the steps outlined in this paper can be reproduced to assess the debate in future pandemics.

As discussed in Section 4.1, the model has high levels of performance on unseen House of Commons speeches. It prioritizes the F1 scores of Class 1 and Class 2 and reliably performs well. Furthermore, reading the speeches which the model has correctly classified illustrates that it appears to pick up on nuanced relationships between attitudes toward COVID-19 restrictions and other political views. The MLC also seems to perform well when a speech includes language arguing both for and against the COVID-19 restrictions. In the majority of cases, it correctly identified which argument was the speaker's primary argument.

Section 4.2 goes beyond testing the model on the test set to predict the class of unlabelled speeches and compare these predictions to other data. This is part of the model validation process and generally contributed to my confidence in the model's behaviour. The model's predictions generally coincided with my intuitive expectations of how political speech would vary across Parties, in relation to the changing severity of the pandemic, and in reaction to policy responses. This validation process was not intended to identify minute weaknesses in the model, but instead to offer a macro view of the model's predictions and assess whether there are glaring surprises or errors. For the most part, this process contributed to my confidence in the model and demonstrated how its predictions could be used to test relationships with other data.

The model has several limitations which are useful to understand in order to apply it correctly and best interpret its predictions. First, the model has been trained and tested on U.K. House of Commons speeches alone and as a result it should not be applied to other kinds of speech without retraining the model. Furthermore, it has only been tested on speeches made during 2020-2021 about the COVID-19 pandemic so before it can be used on speeches about another pandemic, its performance would need to be re-tested. A second limitation is regarding its performance on the test set of speeches. Table 4 shows that the model's biggest weakness is its Class 1 specificity. It tends to over-predict Class 1, specifically, incorrectly identifying Class 3 speeches as Class 1. This is true to a lesser degree with Class 2. This limitation is important, but given the primary metrics that the model is being used to develop are specific to Class 1 and Class 2, it is a preferable one than Class 1 speeches and Class 2 speeches being confused for one another. Specificity and recall are very high when we limit the predictions to just Class 1 and Class 2 speeches. Section 4.1 outlines several other potential weaknesses of the model I gathered from reading misclassified speeches. Language about vaccines and increased testing appear to pose a problem for the model, for example, because they are related to COVID-19 policy interventions that have similar objectives to restrictions but they do not actually impose restrictions.

Hypothesis 2 is that there is a relationship between political speech on COVID-19 restrictions and public compliance as measured with Google mobility data. Section 4.3 offers evidence of such a correlation. There are a number of possible explanations for this relationship. One is that political speech can impact public behaviour. This explanation is informed by previous research on what motivates people to comply with rules, especially Tyler and Jackson (2014), Hasseldine et al. (2003), and Jackson and Bradford (2021), which suggest that deterrence plays a minor role and the norms and communication surrounding rules is crucial in determining rates of compliance. It is possible, therefore, that political speech contributes to norms and the "expressive function of the law" and motivates higher rates of compliance.

However, establishing causality between variables like these two is very challenging. It is equally likely that the direction of causation runs from voters' opinions and behaviour to their representatives' language. In all likelihood the relationship works in both directions. It is a limitation of this paper that it cannot contribute more to our understanding of the causal relationship between these variables. A potential avenue for future research is to use the MLC constructed in this analysis or a similar MLC, and its predictions as a new feature, in order to assess the extent to which political speech impacts compliance with pandemic restrictions. This analysis could make use of time series analysis or "lagging" to identify whether political speech occurs temporally prior to differences in compliance or vice-versa. It could also use case studies or natural experiments to add to our understanding on this topic. Another avenue for future research is to dig deeper into what it is about speeches that motivate or detract from compliance. Are certain kinds of appeals more effective than others? And on the other side, are there ways to criticize restrictions without discouraging compliance?

6 Conclusion

This paper outlines a methodology for how to construct a high-performing machine learning classifier with which to distinguish between different attitudes towards COVID-19 restrictions in speeches in the U.K. House of Commons. It validates the best-performing model both against withheld test data and against other related data. It presents evidence that there is a positive correlation between arguments supporting COVID-19 restrictions and rates of compliance with those restrictions in areas represented by the speakers. The model used in this paper is limited in terms of which kinds of speech and which topics it can be applied to, but its performance demonstrates that the approach taken is a viable way of assessing political speech with quantitative text analysis. The paper does not establish a causal relationship between political speech and restriction compliance but it demonstrates a useful methodology and suggests several avenues for further research on how political speech relates to public behaviour like restriction compliance. Future pandemics are inevitable and better understanding this relationship is a worthwhile endeavour. Machine learning techniques like the one used in this paper can make important contributions.

References

Barberá, Pablo et al. (Jan. 2021). "Automated Text Classification of News Articles: A Practical Guide". In: *Political Analysis* 29.1, pp. 19–42. ISSN: 1047-1987. DOI: 10.1017/PAN.2020.8. URL: https://www.cambridge.org/core/journals/political-analysis/article/abs/automated-text-classification-of-news-articles-a-practical-guide/10462DB284B1CD80C0FAE796AD786BC6.

Bengtsson, Linus et al. (Aug. 2011). "Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti". In: *PLoS Medicine* 8.8. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1001083.

Benoit, Kenneth, Audrey Alejandro, et al. (2019). "Text as data: An overview (Chapter 28)". In: *The SAGE Handbook of Research Methods in Political Science and International Relations*.

Benoit, Kenneth, Kohei Watanabe, et al. (Oct. 2018). "quanteda: An R package for the quantitative analysis of textual data". In: *Journal of Open Source Software* 3.30, p. 774. DOI: 10.21105/JOSS.00774.

Briscese, Guglielmo et al. (2020). "NBER Working Paper Series Compliance with COVID-19 Social-Distancing Measures in Italy: The Role of Expectations and Duration". In: URL: http://www.nber.org/papers/w26916.

Bults, Marloes et al. (2015). *Perceptions and behavioral responses of the general public during the 2009 influenza A (H1N1) pandemic: A systematic review*. DOI: 10.1017/dmp.2014.160.

Cairney, Paul and Adam Wellstead (Oct. 2020). "COVID-19: effective policymaking depends on trust in experts, politicians, and the public". In: *Policy Design and Practice*. ISSN: 2574-1292. DOI: 10.1080/25741292.2020.1837466.

Chawla, Nitesh V et al. (2002). "SMOTE: Synthetic Minority Over-sampling Technique". In: *Journal of Artificial Intelligence Research* 16, pp. 321–357.

Chen, Tianqi and Carlos Guestrin (Aug. 2016). "XGBoost". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM. ISBN: 9781450342322. DOI: 10.1145/2939672.2939785.

Cho, Hichang and Jae Shin Lee (2015). "The influence of self-efficacy, subjective norms, and risk perception on behavioral intentions related to the H1N1 flu pandemic: A comparison between Korea and the US". In: *Asian Journal of Social Psychology* 18.4. ISSN: 1467839X. DOI: 10.1111/ajsp.12104.

Clark, Cory et al. (Jan. 2020). "Predictors of COVID-19 voluntary compliance behaviors: An international investigation". In: *Global Transitions* 2, pp. 76–82. ISSN: 2589-7918. DOI: 10.1016/J.GLT.2020.06.003.

Davis, Jesse and Mark Goadrich (2006). "The relationship between Precision-Recall and ROC curves". In: *Proceedings of the 23rd international conference on Machine learning - ICML '06*. New York, New York, USA: ACM Press. ISBN: 1595933832. DOI: 10.1145/1143844.1143874.

Dietterich, Thomas G. (Aug. 2000). "An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization". In: *Machine Learning 2000 40:2* 40.2, pp. 139–157. ISSN: 1573-0565. DOI: 10.1023/A:1007607513941. URL: https://link.springer.com/article/10.1023/A:1007607513941.

Domingos, Pedro and Michael Pazzani (1997). "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss". In: *Machine Learning 1997 29:2* 29.2, pp. 103–130. ISSN: 1573-0565. DOI: 10.1023/A:1007413511361. URL: https://link.springer.com/article/10.1023/A:1007413511361.

Drummond, Chris and Robert C Holte (2003). "C4.5, Class Imbalance, and Cost Sensitivity: Why Under-Sampling beats Over-Sampling". In: *Proceedings of the ICML'03 Workshop on Learning from Imbalanced Datasets*.

Dugan-Knight, Max (2021). *Covid Political Speech. Github Repository.* URL: https://github.com/maxduganknight/covid-political-speech.

Felbo, Bjarke et al. (2017). "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA, USA: Association for Computational Linguistics. DOI: 10.18653/v1/D17-1169.

Google LLC (2020). *Google COVID-19 Community Mobility Reports*. URL: https://www.google.com/covid19/mobility/.

GOV.UK (2011). *2011 Rural-Urban Classification of Local Authorities and other geographies*. URL: https://www.gov.uk/government/statistics/2011-rural-urban-classification-of-local-authority-and-other-higher-level-geographies-for-statistical-purposes.

Graham, Paul (2002). *A Plan for Spam*. URL: http://paulgraham.com/spam.html.

Guo, Xinjian et al. (2008). "On the class imbalance problem". In: *Proceedings - 4th International Conference on Natural Computation, ICNC 2008* 4, pp. 192–201. DOI: 10.1109/ICNC.2008.871.

Hartmann, Jochen et al. (Mar. 2019). "Comparing automated text classification methods". In: *International Journal of Research in Marketing* 36.1, pp. 20–38. ISSN: 0167-8116. DOI: 10.1016/J.IJRESMAR.2018.09.009.

Hasseldine, John et al. (2003). "Framing, gender and tax compliance". In: *Journal of Economic Psychology* 24.4, pp. 517–533. ISSN: 0167-4870. URL: https://EconPapers.repec.org/RePEc:eee:joepsy:v:24:y:2003:i:4:p:517-533.

Jackson, Jonathan and Ben Bradford (May 2021). "Us and Them: On the Motivational Force of Formal and Informal Lockdown Rules". In: *LSE Public Policy Review* 1.4. ISSN: 2633-4046. DOI: 10.31389/lseppr.24.

James, Gareth et al. (2000). "An introduction to Statistical Learning". In: *Current medicinal chemistry* 7.10, pp. 995–1039. DOI: 10.1007/978-1-4614-7138-7.

James, Joby, L. Sandhya, and Ciza Thomas (2013). “Detection of phishing URLs using machine learning techniques”. In: *2013 International Conference on Control Communication and Computing, ICCC 2013*, pp. 304–309. DOI: 10.1109/ICCC.2013.6731669.

Jeffrey, Benjamin et al. (July 2020). “Anonymised and aggregated crowd level mobility data from mobile phones suggests that initial compliance with COVID-19 social distancing interventions was high and geographically consistent across the UK”. In: *Wellcome Open Research 2020 5:170* 5, p. 170. DOI: 10.12688/wellcomeopenres.15997.1. URL: https://wellcomeopenresearch.org/articles/5-170.

Kotsiantis, Sotiris, Dimitris Kanellopoulos, and Panayiotis Pintelas (2005). “Handling imbalanced datasets: A review”. In: *GESTS International Transactions on Computer Science and Engineering* 30.

Kuhn, Max (2019). *The caret Package*. URL: https://topepo.github.io/caret/.

Levi, Margaret, Tom R. Tyler, and Audrey Sacks (Jan. 2013). “The Reasons for Compliance with Law”. In: *Understanding Social Action, Promoting Human Rights*, pp. 70–99. DOI: 10.1093/ACPROF:OSO/9780195371895.003.0004.

Lewis, David D. et al. (2004). *A New Benchmark Collection for Text Categorization Research.* URL: https://www.researchgate.net/publication/220320442_RCV1_A_New_Benchmark_Collection_for_Text_Categorization_Research.

Li, Xiaojiang (Oct. 2020). “Association Between Population Mobility Reductions and New COVID-19 Diagnoses in the United States Along the Urban–Rural Gradient, February–April, 2020”. In: *Preventing Chronic Disease* 17. ISSN: 1545-1151. DOI: 10.5888/PCD17.200241.

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schutze (2008). *Introduction to Information Retrieval*. Cambridge: Cambridge University Press, pp. 117–118. ISBN: 9780511809071. DOI: 10.1017/CBO9780511809071.

Odell, Evan (2020). *Hansard Data 3. Github Repository.* URL: https://github.com/evanodell/hansard-data3.

Office for National Statistics (2021). *Coronavirus (COVID-19) latest insights*. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/.

— (2019). *Open Geography Portal : Ward to Westminster Parliamentary Constituency to Local Authority District (December 2016) Lookup in the United Kingdom*. URL: https://geoportal.statistics.gov.uk/datasets/ward-to-westminster-parliamentary-constituency-to-local-authority-district-december-2016-lookup-in-the-united-kingdom/explore.

Park, Crystal L. et al. (Aug. 2020). "Americans' COVID-19 Stress, Coping, and Adherence to CDC Guidelines". In: *Journal of General Internal Medicine* 35.8. ISSN: 0884-8734. DOI: 10.1007/s11606-020-05898-9.

Paternoster, Raymond (June 1987). "The deterrent effect of the perceived certainty and severity of punishment: A review of the evidence and issues". In: *Justice Quarterly* 4.2. ISSN: 0741-8825. DOI: 10.1080/07418828700089271.

Peak, Corey M et al. (Oct. 2018). "Population mobility reductions associated with travel restrictions during the Ebola epidemic in Sierra Leone: use of mobile phone data". In: *International Journal of Epidemiology* 47.5. ISSN: 0300-5771. DOI: 10.1093/ije/dyy095.

Pebesma, Edzer (June 2021). "Simple Features for R [R package sf version 1.0-1]". In: URL: https://CRAN.R-project.org/package=sf.

Ratanawongsa, N et al. (Jan. 2013). "Communication and medication refill adherence: the Diabetes Study of Northern California". In: *JAMA internal medicine* 173.3, pp. 210–218. ISSN: 2168-6114. DOI: 10.1001/JAMAINTERNMED.2013.1216. URL: https://pubmed.ncbi.nlm.nih.gov/23277199/.

Soper, George A. (1919). "The lessons of the pandemic". In: *Science* 49.1274. ISSN: 00368075. DOI: 10.1126/science.49.1274.501.

Statista (Aug. 2020). *UK: mobile phone ownership 1996-2018*. URL: https://www.statista.com/statistics/289167/mobile-phone-penetration-in-the-uk/.

The Economist (Nov. 2020). *Why Europe's second, less severe lockdowns are working.* URL: https://www.economist.com/graphic-detail/2020/11/28/why-europes-second-less-severe-lockdowns-are-working?utm_campaign=editorial-social&utm_medium=social-organic&utm_source=twitter.

The Institute for Government (2021). *Timeline of UK government coronavirus lockdowns.* URL: https://www.instituteforgovernment.org.uk/charts/uk-government-coronavirus-lockdowns.

Tyler, Tom R. (1990). "Why People Obey the Law". In: *Yale University Press*, p. 299.

Tyler, Tom R. and Jonathan Jackson (Feb. 2014). "Popular legitimacy and the exercise of legal authority: Motivating compliance, cooperation, and engagement". In: *Psychology, Public Policy, and Law* 20.1, pp. 78–95. DOI: 10.1037/A0034514. URL: /record/2013-40953-001.

Wright, Liam, Andrew Steptoe, and Daisy Fancourt (Oct. 2020). "What predicts adherence to COVID-19 government guidelines? Longitudinal analyses of 51,000 UK adults". In: *medRxiv*, p. 2020.10.19.20215376. DOI: 10.1101/2020.10.19.20215376. URL: https://www.medrxiv.org/content/10.1101/2020.10.19.20215376v1%20https://www.medrxiv.org/content/10.1101/2020.10.19.20215376v1.abstract.

Wright, Marvin, Stefan Wager, and Philipp Probst (2020). "A Fast Implementation of Random Forests". In: DOI: 10.1080/10618600.2014.983641. URL: https://doi.org/10.1080/10618600.2014.983641..

Yu, Bei, Stefan Kaufmann, and Daniel Diermeier (July 2008). "Classifying Party Affiliation from Political Speech". In: *Journal of Information Technology & Politics* 5.1. ISSN: 1933-1681. DOI: 10.1080/19331680802149608.

Zajenkowski, Marcin et al. (Nov. 2020). "Who complies with the restrictions to reduce the spread of COVID-19?: Personality and perceptions of the COVID-19 situation". In: *Personality and Individual Differences* 166. ISSN: 01918869. DOI: 10.1016/j.paid.2020.110199.

7 Appendix

7.1 Speech labelling key

Class	Description
1	The speech mentions support for COVID-19 restrictions[4] or includes language that supports individuals following existing restrictions and acting carefully or criticizes government for imposing too few or too lenient restrictions. If the speech mentions both support and criticism, classify based on which is mentioned more and whether the thrust of the speech is supportive or critical of restrictions.
2	The speech criticises or opposes restrictions or is supportive of easing restrictions. If the speech mentions both support and criticism, classify based on which is mentioned more and whether the thrust of the speech is supportive or critical of restrictions. The important part is not bring "critical" of the government, it is that the argument is for fewer or more relaxed restrictions.
3	The speech does not fall into either Class 1 or Class 2. It may be perfectly balanced between the classes or it may be related to another aspect of the pandemic other than restrictions.

[4] My definition of COVID-19 restrictions: rules, guidelines, or laws specifically intended to prevent the spread of the COVID-19 virus and which have some restrictive component on people. This includes but is not limited to: stay-at-home orders, national lockdowns, closing of businesses, closing of schools, quarantine orders on incoming travellers, and rules mandating the wearing of masks. It does not include efforts to create a vaccine or offering increased testing because these proposals do not restrict people.

7.2 Most important terms in the final Random Forest classifier as measured by their contribution to overall reduction in Gini Impurity

Top features		
virus	home	community
pandemic	disease	tracing
covid-19	measures	economy
protect	weeks	cases
must	course	test
support	bill	first
nhs	clear	learned
lockdown	hon	working
right	ppe	billion
including	transmission	year
people	spread	outbreak
testing	get	time
going	million	stop
prime	take	recent
can	friend	isolate

7.3 Speeches

7.3.1 Andrea Leadsom, 2/11/2020 (South Northamptonshire) (Con)

"The lockdown since March has been devastating for many people and only very reluctantly will I be supporting the latest lockdown measures when they come to the House on Wednesday. Does my right hon. Friend agree that the real problem is for people's mental health, whether it is elderly people who are in care homes or who are desperately missing their families; business people who are seeing their life's efforts ruined around them; or, of course, families with very young children who are isolated and, frankly, miserable? Will he do everything possible to make sure that this lockdown is a compassionate one and that those who are vulnerable and who have mental health problems will be supported through it?"

7.3.2 Ian Blackford, 25/03/2020 (Ross, Skye and Lochaber) (SNP)

"As of Monday, more than 3,300 inquiries have been made in Scotland about NHS staff seeking to return to work to help us defeat coronavirus. Those people and all those already working tirelessly in our NHS are our heroes. Every last one of them, from consultants to cleaners, carers to nurses, drivers to maintenance workers, GPs to paramedics, are performing vital work to save the lives of others. When the crisis is over, we in this House will need to find some way to honour those amazing heroes, but there is one way that the public can honour and support our NHS staff now: by staying at home. Staying at home and adhering to social distancing will save lives, protect our health and social care services and begin to flatten the curve. We can avoid unnecessary deaths, but only if we all act together. Does the Prime Minister agree that we owe it to everyone in our NHS and those willing to return for non-essential workers to stay at home?"

7.3.3 Andrew Gwynne, 16/07/2020 (Denton and Reddish) (Lab)

"I thank the Secretary of State for his statement. We know that distinct areas of the country are seeing local rises in the number of cases, so can he explain what urgent steps the Government are taking to increase testing in those areas? With his indulgence, as someone who is on week 17 of long COVID-19 viral fatigue, may I also ask the Secretary of State what additional resources he is committing to NHS support services for those who are, bluntly, struggling to recover from the virus?"

7.3.4 Valerie Vaz, 25/03/2020 (Walsall South) (Lab)

"I thank the Leader of the House for his statement. Let me start by wishing Prince Charles a speedy recovery. I know that he has tested positive for coronavirus, and that our gracious sovereign is also in self-isolation. It was good that Prince Charles was able to have a test. Many of our front-line staff do not have that test. The Prime Minister said earlier that he wants to protect the NHS. The staff need protecting and they deserve our gratitude, so will the Leader of the House do all that he can to ensure that tests are available for them? The Leader of the House will know that Labour Front Benchers and those of the other Opposition parties are working constructively together, and I hope that will continue when we go into recess. Many of the fiscal measures have come through because our constituents, some of whom are absolutely desperate, have contacted us to ensure that we put their cases forward. I am slightly concerned about the Leader of the House's caveat on 21 April. I know he will do all that he can to ensure that Parliament returns on 21 April, and we know that we are able to operate, albeit with a skeleton staff. May I ask him about voting, because that is another area that hon. Members have concerns about? I am sure that he would be the first to agree that we need to hold the Government to account. We found new ways of voting during the Brexit debate, and therefore I wonder whether negotiations could continue through the usual channels, because clearly voting arrangements must reflect the wishes of the House. I have raised with him the possibility of questions. We know that questions are not answered during recess — and in the light of your statement, Mr Speaker, there is no way that the civil service can cope with 60 questions at a time, and we do not ask for that — but given the unusual times, will the Leader of the House look at ways in which urgent questions can be answered, whether that is through questions or more MPs' hotlines? May I ask the Leader of the House about the Boundary Commission report, which was published as a written statement yesterday and is to be decided by Order in Council? We both know that it is not for the gracious sovereign to be involved in a political decision, so will he ensure that any oral statement comes back to the House so that the House can decide on that?I am tempted to say that I have received an email from the Leader of the Opposition, but I want to pay tribute to him and thank him for all his work, and particularly his family and his staff. They have worked very hard. My right hon. Friend must have done something right, because he has seen off two Prime Ministers.Finally, I want to thank everyone here — the reduced staff who have enabled us to carry on working here and to carry on business — and I want to wish every single hon. and right hon. Member and their families well. I hope that they will be healthy and safe."

7.3.5 Charlotte Nichols, 06/01/21 (Warrington North) (Lab)

"Special schools were not mentioned in the Prime Minister's statement, but they will remain open over the course of lockdown. Will he please advise the House what advice and support they have received to stay open safely for the often vulnerable young people who need them, and whether special educational needs school staff, students and their parents will be given priority access to the vaccine to keep them safe?"

7.3.6 Priti Patel, 08/06/2020 (Witham) (Con)

"The right hon. Lady is right that we should absolutely reflect on the majority who have protested peacefully, and I commend the young people in particular. Online protests are much safer when it comes to the health epidemic that we are enduring right now. Importantly, the voices of those who protested peacefully have in effect been subverted through the violence that we saw this weekend."

7.3.7 Bob Neill, 04/11/2020 (Bromley and Chislehurst) (Con)

"It is a profound moment in which we are being asked knowingly to restrict the civil liberties of our fellow citizens to an unprecedented degree in peacetime, and knowingly and deliberately to harm the economic welfare and, in some cases the personal welfare of our fellow citizens, because lockdowns have consequences and do damage. In deciding whether that can ever be acceptable in a country that believes in the rule of law, it is important to consider whether such measures are necessary, proportionate and supported by evidence. I accept that the covid pandemic is an emergency of a kind that can make such draconian measures necessary. I regret to say, however, that I do not believe that the measures set out in the regulations are either proportionate or based on the evidence.I do not doubt the good intentions of the Secretary of State and the Government, but the details of the measures go beyond those that are appropriate to achieve the objective that is set out. We could refer perhaps to the briefing from the Bingham Centre for the Rule of Law and its reference to the late Sir John Laws who suggested doing the minimum that is necessary to achieve the objective with the minimum intrusion on civil liberties. I am afraid some of the measures here go beyond that. There is no scientific basis for the banning of non-contact outdoor sports. There is no scientific basis for treating grassroots football and community sport differently from elite sports. There is no scientific basis for stopping and, indeed, criminalising people of faith joining in collective worship when they do so in a safe fashion, forgoing the right to join in communal hymn singing or music to limit the risk of transmission. That goes beyond that which is proportional.There is no economic impact assessment, but as to the disbenefits to businesses, I have seen family businesses of 20 years' standing already go under in my constituency. I cannot vote to support that without clear evidence as to why it is necessary, the extent to which it is likely to continue and what the plan is to come out the other side in good order.With a heavy heart, I cannot support the Government in the Lobby today. These measures are not amendable, but I would have been prepared to look at a more limited or proportionate form of regulations, An example of the short notice that we have had to consider these measures and the poor drafting of them is that people are allowed to go to an estate agent, but they cannot go to a solicitor. But the documentation that people will need to get a mortgage and to move house will frequently need to be witnessed in person by a solicitor. These are poorly drafted regulations, and that is only one of many examples. That is why I cannot support them."

7.3.8 Angela Eagle, 14/10/2020 (Wallasey) (Lab)

"The Government have already conceded that fighting the spread of this dangerous covid-SARS virus in our country requires extraordinary levels of state action and support, but now, just as the fight is intensifying, it is clear that they have lost their nerve. We are not only battling this deadly virus; the Prime Minister is fighting his libertarian instincts and the right-wing ideologues in his party. They are opposed to the collective state action that is necessary to save lives and mitigate the damage from the pandemic. The delay that this fight caused in March left us with a double whammy of the highest per-capita death toll in Europe on top of the largest economic hit in the G7, and now, this unforgivable dereliction of duty looks like it is happening again. As the Prime Minister dithers, the virus spreads. His failure to take timely and firm action will cost more lives and wreak more damage on our economy. As he courts his mutinous Back Benchers and abandons the science to keep them sweet, all the warning signs are flashing red again. He is behind the curve and he knows it, and since the SAGE minutes were published on Monday night, we all know it, too.The Government have lost the trust that they need to lead the fight against this deadly threat. Their partisan, high-handed behaviour has made it worse, excluding Parliament completely. There are constant briefings to the media, and an obsession with outsourcing and centralisation has caused the failure of Test and Trace and the scandal of PPE supplier contracts to Tory donors. And:"We will do whatever it takes" — has now turned into the inadequate furlough-lite proposals that the Chancellor has recently come up with. Just as the virus returns, he has packed up the safety net.For my constituents in Wallasey, who are now in tier 3 and facing a local lockdown, vital support disappears at the end of the month. In Wirral, 31,000 people are still on furlough and it will disappear at the end of the month, just as the virus comes roaring back. What replaces it is completely inadequate, as the Chancellor knows only too well, and those who are losing their jobs or their business do not want a lecture from him about how much he has already spent. Those who are excluded completely from this support in the first place — the freelancers, some of the self-employed — do not want that lecture either. They want a Government who will recognise the hardship that the pandemic has caused and be there to help. The least that the Government could have done was to repurpose the £40 million in unspent support allocated to the Liverpool city region, which is now in tier 3, to support local businesses, but again today the Chancellor has refused even that modest request.Those forced to self-isolate to stop the spread of the virus need the support to do so and not to have to choose between feeding their family and obeying the rules. Wirral Council, which has been at the forefront of the fight against the virus, has not been reimbursed for what this has cost and, like many other local authorities, it is teetering on the verge of bankruptcy.So what do we need? We need an increase in generosity of the furlough-lite scheme. It has to pay more to those whose jobs are affected. We need wider eligibility; it has to go to businesses that are affected, not only those that close. We need to include the excluded, which means freelancers and the self-employed, and we need to pay adequate sick pay for those forced to isolate. If we do not do that, the virus will roar back, and the economic cost will, in the end, be far greater and the cost in lives will be unbearable."