

ANALYSING THE TREND OF ISLAMOPHOBIA IN BLOG COMMUNITIES USING MACHINE LEARNING AND TREND ANALYSIS

Complete Research

Tiffany Massey, EY, London, tiffmassey@hotmail.co.uk

Chintan Amrit, Amsterdam Business School, P.O. Box 15953, 1001 NL Amsterdam, c.amrit@uva.nl

Guido van Capelleveen, University of Twente, P.O. Box 217, 7500 AE, The Netherlands, g.c.vancapelleveen@utwente.nl

Abstract

Blogs have been instrumental in shaping public opinion, and constitute an important component of the burgeoning Social Media space. However, researchers have not considered the impact of blog posts and the comments on blog posts to understand public opinion on different topics. This article analyses the trend of Islamophobia in certain blog communities in UK, using public opinion from blog comments taken from a range of political blogs. A proportion of the blog comments were labelled manually, before being used to train an algorithm to label the remaining comments. The algorithms gave varying results, the best being a Bagging algorithm – which is an ensemble algorithm that combines multiple algorithms. After labelling these comments, we answered our research question: Can one identify the trend in Islamophobia by analysing blog comments and if it is related to terror attacks in a particular country? We concluded that there has not been a rise in Islamophobia, but that terror attacks in the UK and abroad caused spikes in anti-Islam comments on the blogs. The main contribution of our research is in demonstrating a method for analysing blog comments to identify the trend in Islamophobia in the blog communities of a country.

Keywords: Islamophobia, Machine learning, Blogs, Trend analysis

1 Introduction

The number of hate crimes against Muslims in the UK has been on the rise, in 2014 the BBC reported a 65% rise from 2013. Terror attacks have also caused spikes in anti-muslim attacks, for example, the days following the London Bridge attack in 2017 saw the number rise fivefold (Dodd and Marsh, 2017), and the Manchester bombing caused Muslim leaders in the city to voice their concerns regarding the rise in hate crimes (Grierson and Booth, 2017). From these reports, and many others (Sky News, 2017; Batchelor, 2017; Mortimer, 2017), there seems to be a rise of Islamophobia in the UK.

Social media has initiated a transformation of business and governments, to increasingly analyse their social media coverage, and learn from trends about users' opinions on social media, related to their company or on a specific subject (Aral et al., 2013). The digital trace of internet users, in particular, the issues in the blogosphere, provide a rich analysis of the behavior of individuals related to sociological phenomena (Lazer et al., 2009). Furthermore, the big data present in online social networks allows the study of patterns of influence and/or information propagation of online social networks based on computational social science (Abbasi, 2016). Silva et al. (2009) explore the social dynamics of blog communities through the theoretical lens of communities of of practice. They describe four primary practices that bring about cohesion in blog communities.

This research aims to determine whether there has also been a rise in Islamophobia in certain blog communities, using comments on blog posts from July 2014 until June 2017. This time scale was chosen because the EU Referendum was announced early in 2016, so to compare data a year and a half on each side of 2016 seemed an appropriate time period. Although many researchers have analysed the trend of Islamophobia over the last twenty years, as far as we are aware, this research is the first to analyse blog comments in the context of Islamophobia. We decided to analyse blog comments rather than blog posts, as blog comments would show a larger range of opinion of the blog community (as done by Silva et al, 2009). It was also noted that a blogger may not openly express Islamophobic views in order to remain politically correct and to not damage their reputation, whereas comment posters can remain anonymous when posting. Our main research question is: *Can one identify the trend in Islamophobia in blog communities by analysing blog comments and if this trend is related to terror attacks in a particular country?*

We analysed blog comments from a range of UK political blogs and performed trend analysis to answer our research question. The final conclusions made from this analysis will not represent the opinion of the UK as a whole, but will still show the trend for the comment posters from the six popular blogs used.

We specifically focus on certain blogs in the UK and use supervised machine learning techniques that will be accepted when an appropriate accuracy is reached. Once this has been investigated, a range of statistical analysis techniques will be used to shed light on whether there has been a rise in the rate of Islamophobia in the UK since 2014. TellMAMA ¹ reported that anti-Muslim attacks increased by 326% in 2015 (Sherwood, 2016) and it would be interesting to see if this is reflected in the form of blog comments. As noted in our research question we are interested to find out if (i) certain events, such as terror attacks in the UK and in other countries are related to the proportion of negative comments, and (ii) if our method of analysing blog comments can help in identifying the trend in Islamophobia. As mentioned previously, the number of anti-Muslim hate crime rises after terror attacks, we are curious to see if this is mirrored in our analysis of blog comments. We are also interested in finding out if Islamophobic comments more likely to be seen in right, centre or left wing blogs. In the 2017 UK General Election, UKIP campaigned to ban the burqa in the UK (Elgot, 2017). We would like to investigate if this is mirrored in the opinion of right

¹ A national project which records and measures anti-Muslim incidents in the United Kingdom

wing comment posters. Vanparys et al. (2013) noticed a difference in trend in right, centre and left wing blogs in different countries when analysing opinion of Islam. We would like to investigate if we identify similar trends in the blog comments. Our main contribution of this paper is in demonstrating our method of analysing blog comments to identify the trend in Islamophobia.

The rest of the paper will be organised as follows. Section 2 of this research will review the literature available regarding sentiment analysis, machine learning and trend analysis related to Islamophobia. Section 3 will then explain the methodology used for each step of the analysis. Section 4 will show the results of the machine learning methods used, the trend analysis and the statistical tests. Section 5 will discuss the results and give explanations for them, and Section 6 will conclude the research and explore possible areas of further work.

2 Literature Background

In this section of the literature review we focus on how the data can be labelled initially, as many researchers use already labelled data and then use supervised machine learning - to determine the sentiment of the textual data. Similar to the research in this article, Vanparys et al. (2013) analysed the opinion of Islam in six EU countries from 1999 to 2009. The data used is from the EURISLAM project, it was a database of political claims with regard to Muslims and Islam in Europe. Vanparys et al. (2013) used a very simple scoring system; each claim which showed worsening rights for Muslims received a (-1) score, an improvement received a (+1) score and neutral claims received a (0). Vanparys et al. (2013) did not state how they decided which claims were positive, negative or neutral, which could mean bias in the decision making. It was also not said how they decided to categorise the data, as there is often overlap between topics – so some clarification on this would help to increase the legitimacy of their research.

Miller et al. (2016) analysed Islamophobia using tweets from March 2016 to July 2016. They manually determined the sentiment of tweets and fed this through their own algorithm. The algorithm spots patterns in the data and looks for statistical correlations between the language used and class assigned. They used both unigrams and bigrams to determine which are good indicators of which category a tweet should belong to. They had a recall score of 80% and a precision score of 80%, which is a satisfactory result for text classification machine learning. However, because none of the algorithms were 100%, some of the data will have been misclassified and this means that the results are not 100% reliable.

Agrawal and Sureka (2014) endeavoured to classify hate and extremism promoting tweets using machine learning. A semi-supervised learning approach was taken because annotating tweets is arduous and time-consuming. Tweets were identified using hashtags that related to the research, such as #Terrorism, #Islamophobia and #Extremism. The training set of tweets were manually analysed. Agrawal and Sureka (2014) conducted experiment using, firstly, a k-Nearest Neighbour (k-NN) algorithm, and then a Support Vector Machines (SVM) algorithm. Two public datasets of tweets were used from May 2011 and June 2011. Once the algorithms had been run on the data, the results were validated by four graduates manually labelling each tweet. The k-NN algorithm achieved an accuracy of 90% and the SVM algorithm had an accuracy of 97%. This is an interesting way to check accuracy of results, if the graduates had already seen how the classifiers labelled the tweets then they may have been biased when labelling themselves, however the researchers do not confirm whether or not the classifier results were shown.

Trend analysis is a useful way to view data regarding opinion. In this research the proportion of negative comments will be plotted and conclusions will be drawn. Many others have used trend analysis to view the rise in Islamophobia.

Vanparys et al. (2013) plotted the data from their research as a Time Series graph, peaks in “mentions” were observed at significant events such as 9/11 and the London bombing in 2005. When the data was split into claims for each country the peaks were noted in different areas, as some events had a larger effect on the country. As part of their analysis, Vanparys et al. (2013) also split the data into four categories. These were: statements about minority rights and integration, racist or Islamophobic comments, political and Islamic extremism and violence, and an “other” category which relates to asylum and crimes committed. When analysing these four categories it was seen that the minority rights and integration category did not show the 9/11 peak that the others portrayed. This led the researchers to conclude that the splitting of the data was a major step, because of the differences between the categories. Vanparys et al. (2013) also undertook some statistical analysis as part of the research, including intervention analysis. This analysis evaluates the impact of particular events on a time series. They found that the significant events caused an abrupt but short change in the data, rather than a long-term change. It was concluded that their research did not show a heightening in Islamophobia, which contradicts general opinion.

Using four popular UK newspapers, Bleich et al. (2015) analysed the media portrayal of minorities. They used LexisNexis to extract headlines from 2001 to 2012, headlines were extracted if they contained “Muslim”, “Moslem”, “Islam”, “Jew”, “Judai” and “Christian”. The headlines were categorised into “Victim”, “Beneficial” and “Problem”, it was not mentioned how the sentiment of the headlines were determined. The net tone regarding Muslims from 2001 and 2012 was calculated and plotted. Bleich et al. (2015) found that from 2003 — 2008 the net tone was negative, however when each headline was weighted in proportion to newspaper circulation the number of years in which the overall tone was negative increased to include both 2010 and 2012. This does not imply a rise of Islamophobia in the UK, but shows that the view has remained negative for many years.

Bleich et al. analysed New York Time headlines in their 2016 article (Bleich et al. 2016), where they used headlines from 1985 — 2013. Similarly to their 2015 work, Bleich et al. calculated the net tone of the newspaper headlines and found that newspaper headlines have not become more negative since 9/11. This contradicts the general opinion that media portrayal of Muslims and Islam have become worse. This may be due to the fact that they chose to only used headlines from one newspaper, which won’t necessarily represent the trend of all newspapers in the USA.

Miller et al. (2016) plotted a trend graph for July 2016, in which they identified over 215,000 tweets sent in English as being anti-Islamic – this is on average 289 per hour. They observed peaks and tried to work out what had caused them, discovering that it was the terrorist attacks in Nice that cause the spikes.

3 Research Design and Methodology

The internet has a huge amount of blogs and blogging platforms for people to put forward their view. Due to the nature of this research, various blogs used were chosen very specifically using Purposive sampling (Vehovar et al. 2016). To get an idea of the top UK political blogs, four different top blog lists were used. These were: Vuelio’s Political Blogs UK Top 10 (Williams, 2017), Feedspot’s Top 25 UK Political Blogs & Websites on Web (Feedspot, 2017), The Guardian’s Political blogs (Helm, 2010) and Total Politics Top 50 UK Political Blogs (Total Politics, 2010).

The data for this research was collected in 2017. Once the top political blogs were extracted from the lists, each blog was viewed and it was determined whether it was suitable for this research. For example, one of the most popular political blogs is Guido Fawkes (2017). Unfortunately, although very popular, Guido Fawkes is seen as more of a tabloid blog than an informative blog. Other blogs such as Political Scrapbook (2017), did not have enough posts and too few comments to be used for this research. Conservative Home

(2017) had a lot of comments, and would have been a good blog to use for this research, however the comments section of blogs is removed each year. Therefore only comments from the year 2017 would have been available for this research.

The final six blogs that were used are a mixture of right, left and centre wing political views and are a mixture of personal blogs and blogs with multiple writers. It is interesting to have this mixture, especially from the blogs which have multiple contributors as this means that each blog post has a different opinion and may cause a different reaction to another authors post on the same subject.

- **Centre Blogs:**

- **Craig Murray's Blog:** Craig Murray is an author, broadcaster and human rights activist (Murray, n.d.) who was a Liberal Democrat from his University years until 2011 when he became a member of the Scottish Nationalist Party (Murray, 2011). The comments on Craig Murray's blogs are moderated, however most comments are posted. This blog is classed as "centre" for this research, due to Craig Murray's political affiliation.
- **Independent Voices:** The Independent Voices is a section of the Independent news website which allows authors to post their own view. The Independent is classed as a liberal newspaper, and therefore for this research it is defined as "centre".

- **Right Blogs:**

- **Tim Worstall's Blog:** Tim Worstall is a writer who has had articles in The Times and the Daily Telegraph, among other papers. He is a known supporter of UKIP and stood as a candidate for London in the European Parliament election in 2009 (Worstall, 2013). Therefore, for this research this blog is classed as a "right" wing blog.
- **Peter Hitchens Blog:** Peter Hitchens writes columns for the Mail on Sunday and was a foreign correspondent. He was previously a Conservative party member, however he no longer supports them due to his opinion of how the party is run. His column for the Mail on Sunday is classed as "right" wing blog.

- **Left Blogs:**

- **Left Foot Forward:** Left Foot Forward was started by Will Straw, who is a labour politician, in 2009. The blog is a "political blog for progressives" (Feedspot, 2017), which doesn't promote particular parties, but rather progressive goals. In this research this blog is classed as "left" wing.
- **Guardian Comment is Free:** The Guardian comment is free section has various authors who post their opinions about a variety of subjects. These are the only blog posts on the Guardian that allow comments, and comments are only open for a few days before being closed. Similar to Craig Murray's blog, the comments are moderated and a proportion had been removed at the time of this research. The Guardian is known to be a left wing newspaper, and therefore for this research it is classed as a "left" wing blog.

The blog posts from each blog that were chosen to be scraped were those that related to Islam or Muslims in any way. This included posts about Syrian refugees and about Muslim issues in other countries as well as posts specifically about terror attacks in the UK. The blogs were scraped using Scrapy (a Python framework), except for the comments from both the Guardian and the Independent that were scraped manually. In this research each blog was scraped using an independent spider, which enabled the researcher to scrape comments at intervals instead of running the code for an extended period of time.

Scraping the 266 blog posts, resulted in thousands of comments, many of which did not relate to this research. In order to reduce the large amount of comments, comments were filtered based on whether or not they contained any of the following words: *Islam, Muslim, Moslem, Jihad, ISIS, Terror, Bomb* or

Extremist. These words were chosen because this research is specifically researching negative opinion on Islam, and these words are the most basic linked to Islam. We labelled the training set manually. The comments were cut down to one or two sentences based on whether or not the sentence related to the topic. The comments were labelled as 'negative', 'non-negative', as the main aim is to measure the change in negative comments. The earlier mentioned research done by Miller et al. (2016) received a lot of backlash due to the lack of definition of Islamophobia used in their work. Miller (2016) was able to categorise the types of tweets they considered to be Islamophobic. This research categorised negative comments as those which:

- stated, or suggested, that Muslims should all leave the UK (for example "*The only solution is a Britain without a single Muslim. Walls work. We need to return to a much more homogeneous country.*")
- were extremely negative about Islam, or stated Islam should be banned from the UK (for example "*The problem is Islam, not the packages it is in. The West must restrict Islam, as it is hate speech and incites murder*")
- categorised all Muslims as terrorists (for example "*Every Muslim – EVERY MUSLIM – is a potential terrorist. You have no way of knowing which will act soon. Salman Abedi didn't kill Saffie Rousso, Islam did. Until you understand that, you are doomed.*")

To reduce bias, a sample of 65 of the final classified dataset was distributed among three individuals for them to classify themselves. To test for internal consistency of each item, a Cronbach's alpha test was performed in SPSS. The Cronbach's alpha statistic was 88.4% across the four samples. This is a very good score, because generally over 70% is accepted.

3.1 Methodology

We used machine learning to classify the comments. For this we used R and the following packages: the *tm* package: for text mining, the *e1071* package: for Naïve Bayes algorithm, and the *RTextTools* package: for automatic text classification using machine learning. To undertake the trend analysis, R was used to plot the time series of the proportion of negative and to plot the decomposition. From this we determined whether or not there has been an increase in Islamophobia in the blog comments used in this research. To understand if Islamophobic comments are more likely to be seen in some blogs than others, a Kruskal-Wallis test was undertaken in SPSS. In this test the "class" of the comment is the dependent variable, and will be binary, where 1 is "Negative" and 0 is "Non-Negative". The independent variable will be the "Political Affiliation" of the blog, either "Centre", "Left" or "Right". These are the first two assumptions of the Kruskal-Wallis test, the third is that the observations are independent. This research did not scrape the name of the comment poster, and therefore we will assume that the same person has not posted more than once, and in this way we can assume independence.

In order to label the unlabelled comments, a dataset of the 400 manually labelled comments were used to teach a classifier. Firstly, the dataset was read into R as a csv file. The order of the data was randomised to ensure that the training and test sets had a random selection of negative and non-negative comments. This dataset was saved and used in all the algorithms onwards. A corpus of words was then created using the Corpus function from the *tm* package. The data was then preprocessed and the following standard text pre-processing steps were applied: (i) Punctuation was all removed to reduce each comment to purely words, (ii) Numbers were all removed as these does not aid the algorithms in any way, (iii) Stop words were removed (words such as "and", "the", etc.), (iv) Whitespace was removed to ensure the data is cleaner, and (v) All words were converted to lowercase. After the corpus was created, it was converted

Classifier	Naive Bayes	SVM	Boosting	MAXENT	CART	Bagging	Random Forest
Precision	0.51	0.62	0.56	0.60	0.63	0.64	0.56
Recall	0.69	0.57	0.81	0.62	0.67	0.69	0.71
F1	0.59	0.54	0.65	0.61	0.65	0.66	0.69
Accuracy	0.52	0.55	0.56	0.60	0.65	0.66	0.66

Table 1: Results of running 10 fold cross validation on the data set

into a Document Term Matrix. This is a matrix which counts the frequency of each word in a comment.

4 Results

We used 10 fold cross validation for the SVM, Naive Bayes, CART, Random Forest, Bagging, Boosting and Maximum Entropy implementations in R. We tried all algorithms with and without stemming. We found that stemming did not improve the performance of the algorithms in our case.

The accuracies of the machine learning analysis can be seen in Table 1. We found that the Bagging and Random Forest algorithms performed similarly and Bagging was a little more consistent for this dataset. So we used the Bagging algorithm for further analysing our data using trend analysis. Figure 1 shows the trend line of the proportion of negative comments over time. This was plotted in R using the *ts* function. The value for August 2016 had to be imputed, because there were no comments at all for that month. The average of the four months either side of it was taken, as this seemed to be the most appropriate way to impute it.

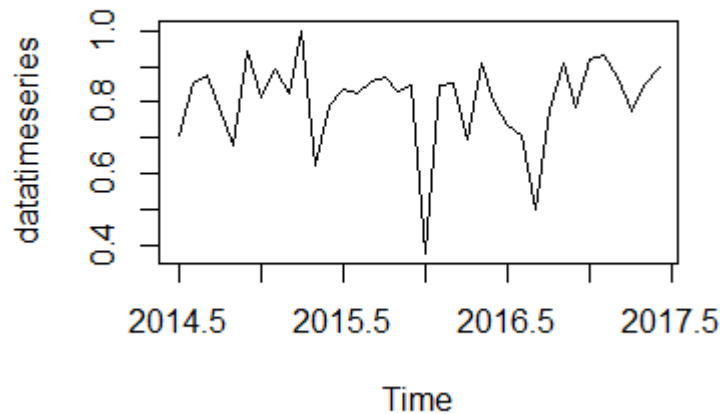


Figure 1: Proportion of Negative Comments over time

It can be seen that there hasn't been a rise in the proportion of negative comments, but to see the trend we also plotted a decomposition of the time series, which can be seen in Figure 2. From the decomposition graph, it can be seen that the trend has decreased. As expected, there is no seasonality in the data – this can be seen by the way the randomisation completely counteracts the seasonality in some parts. To plot Figure 3, a pivot table was created from the final dataset and only the negative comments were selected to be shown. From the graph we can see that the negative comments peaked at multiple terror attacks, both in the UK and elsewhere, including the Charlie Hebdo shooting in Paris and the Manchester Arena

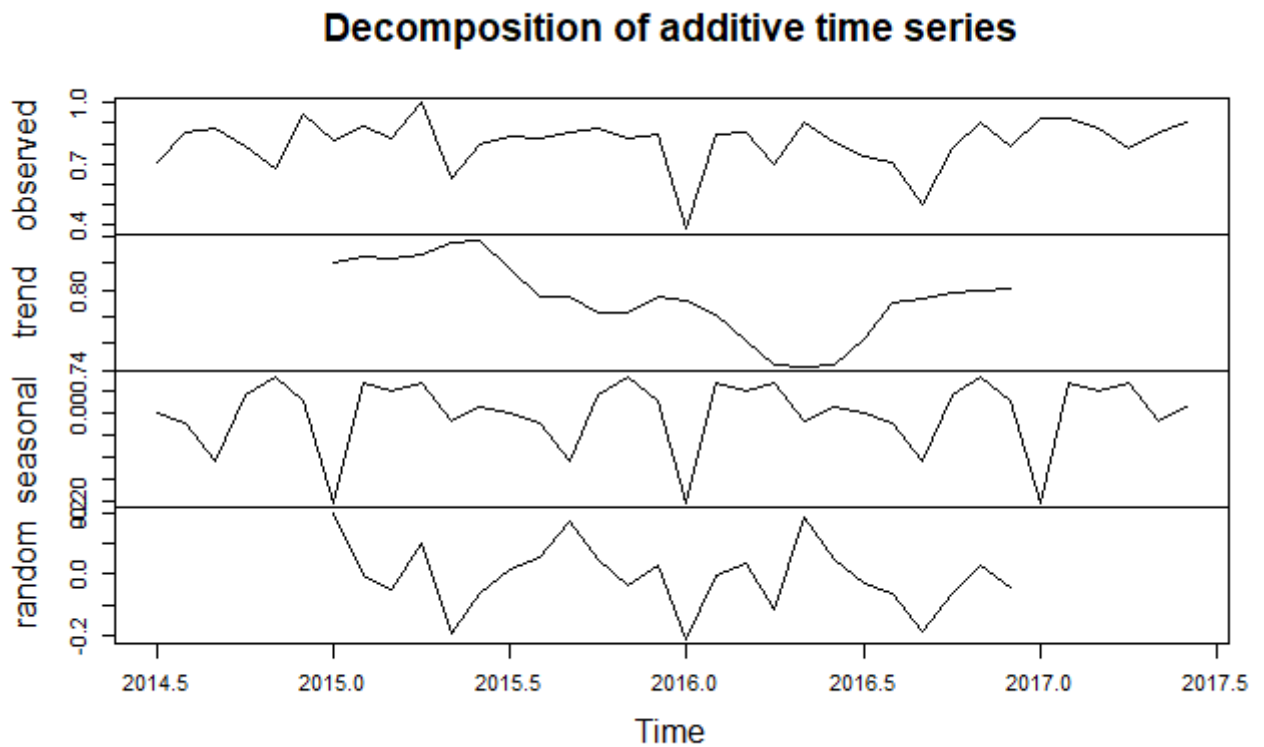


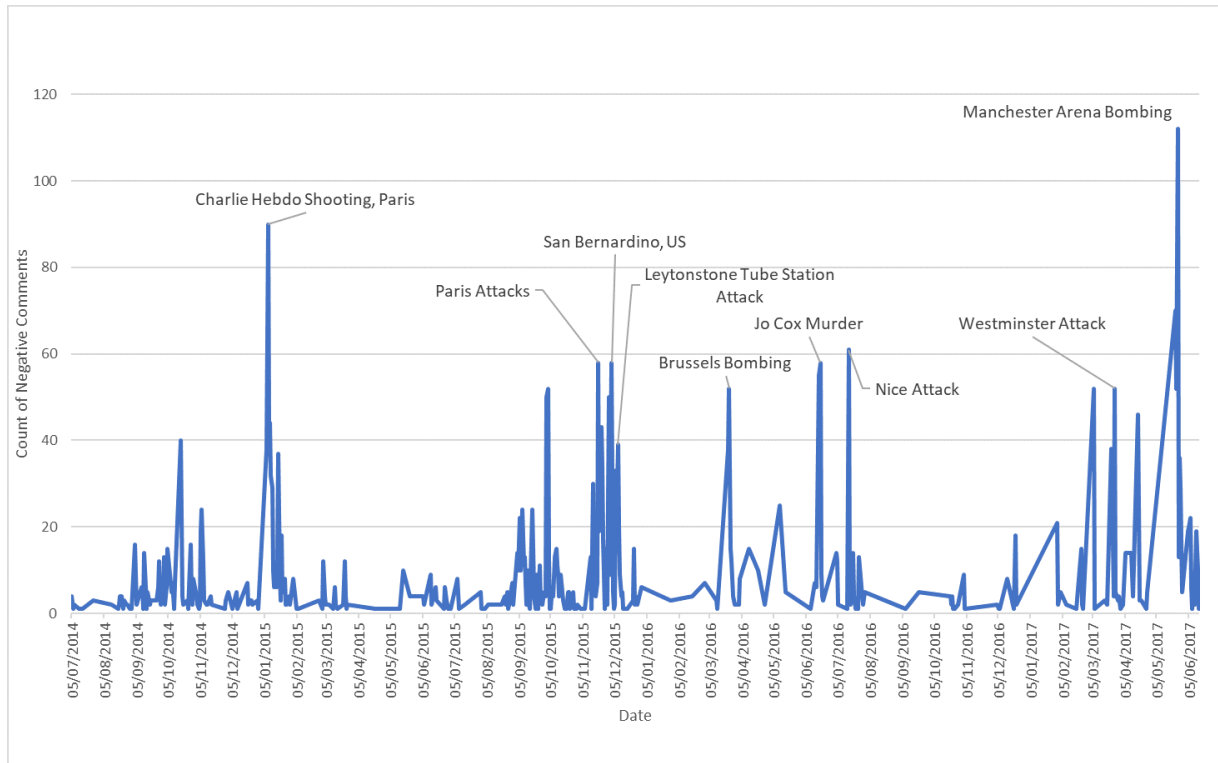
Figure 2: Decomposition of Time Series

bombing. Each peak has been labelled with the event that likely caused it, these will be reviewed and explained in the next chapter.

5 Discussion

We were able to get an accuracy of around 66% using Machine Learning methods in R to classify the comments. This result was achieved in both the Bagging algorithm and Random Forest algorithm. The downside to the Bagging algorithm was that it took a lot longer to run than the other algorithms. In the case where the Bagging algorithm was not the best, the Boosting algorithm often gave the best result. Quinlan (2006) found that Boosting performs better than Bagging, so even though in this research this was not the case – the Boosting algorithm still performed better than the non-ensemble algorithms. The Naive Bayes algorithm on the other hand had the worst results, which could be because the algorithm assumes independence between words, when in practice this is often not the case. Words may be more likely to appear in a negative comment than a non-negative one, and vice versa. Using stemming did not improve the results of any of the classifiers, even though it should have helped to teach the algorithms. It is likely that the inability to reach a higher accuracy than 66% is due to the data itself. Text classification is a difficult task, especially when the text shows opinion. Comment posters can be very sarcastic, which a text classifier probably would not pick up on. To improve the results the researchers considered manually labelling more comments so that a larger training set could be used, however this is time consuming and the results may not have been improved.

From graphs 1 and 2 it can be seen that, from our chosen blogs, there has not been an increase in Islamophobia in certain blog communities in UK. This is a surprising result and contradicts the increase in hate crimes against Muslims on social media (e.g., Awan et al. 2016). However, this result is not uncommon -



(a) Trend of Negative Comments

	Charlie Hebdo Shooting	Paris Attacks	San Bernardino	Leytonstone Tube Station Attack	Brussels Bombing	Jo Cox Murder	Nice Attack	Westminster Attack	Manchester Arena Bombing
Year	2015	2015	2015	2015	2016	2016	2016	2017	2017
Date	7 Jan	13 Nov	2 Dec	5 Dec	22 Mar	16 Jun	14 Jul	22 Mar	22 May
Country	FR	FR	US	UK	BE	UK	FR	UK	UK
Killed	12	137	16	0	35	1	87	6	23
Injured	11	413	24	3	340	1	434	49	800+

(b) Casualties of the terror attack

Figure 3: Analysis of the Islamophobia trend

Vanparys et al. (2013) saw that terror attacks did cause an abrupt change in opinion but that this did not have a lasting effect. As has been mentioned, these results are not representative of the whole of the UK, however it is still interesting that there has been no increase, but that the decomposition of the results actually shows a slight decrease in Islamophobic comments on blog posts. In 2017, there have been four terror attacks by Muslim attackers. This has doubled the amount of terror attacks in the UK this decade. It could be hypothesised that, due to the increased number of attacks, the represented population have become more educated to the nature of terror attacks and are aware that those carrying out the attacks do not represent the entire Muslim population. Over time, this may cause a lessening in negative comments about Islam as a whole. This is well demonstrated in the hashtag #YouAintNoMuslimBruv, which trended after an onlooker to the Leytonstone Tube Station Stabbing in 2015 shouted it at the attacker. Muslims and Muslim leaders in the UK are also quick to condemn terror attacks (Mann, 2017), which shows to the British population that this is not an attack from all Muslims.

From Figure 3, it was seen that terror attacks do tend to cause a spike in negative comments and the largest spikes were labelled with probable causes. These will now be explained below.

Charlie Hebdo Shooting, January 2015 – Mass shooting of Charlie Hebdo staff, in which twelve people died and eleven more were injured, following a cartoon of Muhammad in the satirical newspaper. **San Bernardino mass shooting, December 2015** – A couple shot and killed 16, and injured 24 more in a mass shooting and an attempted bombing in San Bernardino. **Paris attacks, November 2015** – Suicide bombers struck outside the Stade de France, followed by mass shootings in restaurants and cafes across the city. **Leytonstone Tube Station Stabbing, December 2015** – A 29-year old Somalian man attacked three people with a knife at Leytonstone Tube Station whilst shouting “this is for Syria”. **Brussels, March 2016** - Three suicide bombings took place across Brussels in which 35 people were killed and over 300 were injured. **Jo Cox Murder, June 2016** – Jo Cox was a Labour MP who was shot and stabbed by a far-right supporter. **Westminster Attack, March 2017** - Khalid Masood drove a car into pedestrians along Westminster and then stabbed and killed an unarmed police officer. **Manchester, May 2017** - A homemade bomb was detonated in the entrance to the Manchester Arena at the end of an Ariana Grande concert. There are also large observed peaks that do not have an obvious cause:

- 17th October 2014 – This peak is seen a week after a terror attack in Ankara, Turkey. If one of the bloggers mentioned this it might have caused a delayed reaction.
- 1st, 2nd and 3rd October 2015 – There were no terror attacks around these dates, however one could speculate that the spike could have been caused by Russia’s air strikes against ISIS in Syria.
- 6th March 2017 – There were no attacks against European countries at this time, however there were multiple ISIS bombings on this day. It might be that one of the bloggers posted an opinionated piece on this day which caused a spike in negative comments.
- 17th April 2017 – Similar to the spike on the 6th March, there doesn’t seem to be any cause. However, it could have been another delayed spike from the Westminster Attack in late March.

Some terror attacks also caused smaller spikes, including the Berlin Christmas Market attack, in which a truck was deliberately driven into the Christmas Market. This killed 12 and injured 56. This smaller spike was of 18 negative comments. As mentioned previously, there was also a small spike at the beginning of June. This coincides with both the London Bridge attack, in which a van struck pedestrians on London bridge and the three occupants then stabbed people in Borough Market, and the “One Love Manchester” concert mentioned previously, so both may have contributed.

It can be seen from Table 2 that all blogs have a higher proportion of negative comments than positive or neutral comments. This is inline with the bias towards negativity as described by Rozin and Royzman (2001). We also see that the centrist blogs had the highest proportion of negative comments and the left wing blogs had a much lower proportion than the other two. However, most of the comments for all the political affiliations were negative. From the table it can also be seen that the non-negative comments for

the left wing blogs is not much lower than for the others, even though the data set contains less than half the amount of left wing comments than the other two.

Political Affiliation	Negative	Non-Negative	Total	Proportion of Negative
Centre	1534	260	1794	0.855
Left	503	206	709	0.709
Right	1396	256	1652	0.845
Total	3433	722	4155	0.826

Table 2: Proportion of negative comments by political affiliation

To test whether or not there is a difference between the number of negative comments in centrist, right wing and left wing blogs, a Kruskal-Wallis test was undertaken. The result of the test was that the null hypothesis "the distribution of Class is the same across categories of Political" was rejected. Meaning there is a difference in distribution of negative comments across the political affiliation of the blogs. This was because of the left wing blogs, as they have a much lower proportion of negative comments than the centre and right wing blogs.

6 Conclusion and Further Work

The main contribution of this research is in demonstrating the method of analysing blog comments to identify the trend in Islamophobia in the blog communities of UK. From our research we found that Islamophobia, in the six blogs used, has not risen over the last three years. In fact, the data suggests it has fallen since 2015. This is an interesting result, considering the large proportion of negative comments and the public opinion at the current time. As mentioned previously, this could be due to the public being more educated regarding who is behind the attacks, and not blaming the Muslim population as a whole.

We can clarify that certain events definitely do cause spikes in negative comments on blog posts. We observed large spikes at particularly horrific terror attacks, and can conclude that terror attacks change the opinion of the public. However, it is noted that the trend has not gone up, so we can also conclude that the terror attacks cause spikes in negativity rather than a long term change. In this research we were able to get an accuracy of 66% using basic algorithms within R. However, if we had been able to achieve a higher accuracy then we would have less error in our data.

Using a Kruskal-Wallis test we are able to confirm that there is a difference in the proportion of negative comments in some politically affiliated blogs than others. From Table 2 we can conclude that the left wing blogs are the ones that are different to the others. It was expected from the beginning that left wing blogs would have a lower proportion of comments, however it is a surprise that the centrist blogs have the highest proportion of negative comments.

This research has a variety of limitations which will now be discussed. When the final dataset was analysed it could be seen that there is a much smaller amount of comments from the left-wing blogs used. It would have been better if the research had had a more equal amount of blogs from each political affiliation, however, due to the timescale of this research and the selection of blogs available this was unobtainable.

This research has only touched on six blogs out of hundreds within the UK. To improve results and to be able to make broader conclusions, one would need to delve deeper and find more popular blogs. An issue found by the researchers for this article was that many of the blogs that could have been good for the research did not have many, or any, comments and therefore were not appropriate for this research.

To counteract this problem future work could be to use blog titles, such as Bleich et al. (2013, 2016) did with newspaper headings. This would allow the researcher to use a much larger amount of blog sites and would give much more data. Although the trend of negative comments may not be generalizable, given our relatively small sample size. We think that our method of identifying the sentiment of the blogging community is generalizable.

Also, trend development in Islamophobia was analyzed through the number of absolute observations of phenomena. Such a perspective may be too narrow as Islamophobia appears in different levels of fear, hatred, and hostility toward Islam and Muslims. Another insightful perspective to study trend development in Islamophobia is to investigate the effects of people protecting Muslims, or condemning the attacks against them, for which a message chain (discussion) needs to be analysed.

References

- Abbasi, A., Sarker, S., and Chiang, R.H.L. 2016. Big Data Research in Information Systems: Toward an Inclusive Research Agenda, *Journal of the Association for Information Systems*, Vol. 17(2), Article 3.
- Agrawal, S. and Sureka, A., 2014. Learning to classify hate and extremism promoting tweets. In *Intelligence and Security Informatics Conference (JISIC)* (pp. 320-320).
- Aral, S., Dellarocas, C. and Godes, D., 2013. Introduction to the special issue—social media and business transformation: a framework for research. *Information Systems Research*, 24(1) (pp.3-13).
- Awan, I., 2016. Islamophobia on Social Media: A Qualitative Analysis of the Facebook's Walls of Hate. *International Journal of Cyber Criminology*, 10(1).
- Batchelor, T., 2017. London terror attack: Huge rise in Islamophobic hate crime following Borough Market stabbing, police figures show. [online] Available at: <http://www.independent.co.uk/News/uk/crime/london-bridge-attack-latest-rise-islamophobic-hate-crimes-borough-market-stabbing-terror-police-a7777451.html> [Accessed 21 September 2017].
- Bleich, E., Stonebraker, H., Nisar, H. and Abdelhamid, R., 2015. Media portrayals of minorities: Muslims in British newspaper headlines, 2001–2012. *Journal of Ethnic and Migration Studies*, Vol 41 (pp.942-962).
- Bleich, E., Nisar, H. and Abdelhamid, R., 2016. The effect of terrorist events on media portrayals of Muslims and Muslims: evidence from New York Times headlines, 1985-2013. *Ethnic and Racial Studies*, Vol 39 (pp. 1109-1127).
- Conservative Home, 2017. [online] Available at: <https://www.conservativehome.com/> [Accessed 8 July 2017].
- Dodd, V. and Marsh, S., 2017. Anti-Muslim hate crimes increase fivefold since London Bridge attacks. [online] Available at: <https://www.theguardian.com/uk-news/2017/jun/07/anti-muslim-hate-crimes-increase-fivefold-since-london-bridge-attacks> [Accessed 18 September 2017].
- Elgot, J., 2017. Ukip to campaign to ban burqa and sharia courts, says Paul Nuttall. [online] Available at: <https://www.theguardian.com/politics/2017/apr/23/ukip-to-campaign-to-ban-burka-and-sharia-courts-says-paul-nuttall> [Accessed on 10 September 2017].
- Feedspot, 2017. Top 25 UK Political Blogs & Websites on the Web. [online] Available at: https://blog.feedspot.com/uk_political_blogs/ [Accessed on 5 July 2017].
- Go, A., Bhayani, R. and Huang, L., 2009. Twitter sentiment classification using distant supervision. *CS224N Project Report*, Stanford (pp.12).
- Grierson, J. and Booth, R., 2017. Muslim leaders in Manchester report rise in Islamophobic incidents. [online] Available at: <https://www.theguardian.com/uk-news/2017/may/24/muslim-leaders-in-manchester-report-rise-in-islamophobic-incidents> [Accessed 10 September 2017].
- Guido Fawkes, 2017. [online] Available at: <https://order-order.com/> [Accessed on 7 July 2017].
- Helm, T., 2010. The 10 best political blogs. [online] <https://www.theguardian.com/culture/2010/mar/21/10-best-political-blogs> [Accessed on 6 July 2017].

- Hopkins, D.J and King, G., 2010. A method of automated nonparametric content analysis for social science. *American Journal of Political Science* 54(1) (pp.229-247).
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, and others. 2009. Computational social science. *Science*, 323(5915) (pp.721-723).
- Laerd Statistics, n.d. Kruskal-Wallis H Test using SPSS Statistics. [online] Available at: <https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statis-tics.php> [Accessed 20 September 2017].
- Le, J., 2015. The 10 Algorithms Machine Learning Engineers Need to Know. [online] Available at: <http://www.kdnuggets.com/2016/08/10-algorithms-machine-learning-engineers.html> [Accessed 16 August 2017].
- Mann, T., 2017. Muslim unite to condemn Ariana Grande Manchester terror attack. [online] Available at: <http://metro.co.uk/2017/05/23/muslims-unite-to-condemn-ariana-grande-manchester-terror-attack-6655366/> [Accessed 1 October 2017].
- Miller, C., 2016. Measuring Islamophobia on Twitter. [online] Available at: <https://www.demos.co.uk/blog/measuring-islamophobia-on-twitter> [Accessed 28 August 2017].
- Miller, C., Smith, J. and Dale, J., 2016. Islamophobia on Twitter: March to July 2016. [online] Available at: https://www.demos.co.uk/wp-content/uploads/2016/08/Islamophobia-on-Twitter_-March-to-July-2016-.pdf [Accessed 28 August 2017].
- Mortimer, C., 2017. Islamophobic crimes rose after Westminster attacks, police reveal. [online] Available at: <http://www.independent.co.uk/news/uk/crime/islamophobic-crime-numbers-rise-london-terror-attack-westminster-racism-anti-muslim-a765597-1.html> [Accessed 20 September 2017].
- Murray, C., 2011. The Lonely Liberal. [online] Available at: <https://www.craigmurray.org.uk/archives/2011/09/the-lonely-liberal/> [Accessed 28 August 2017].
- Murray, C., n.d. About Craig Murray. [online] Available at: <https://www.craigmurray.org.uk/about-craig-murray/> [Accessed 28 August 2017].
- Political Scrapbook, 2017. [online] Available at: <https://politicalscrapbook.net/> [Accessed on 10 July 2017].
- Quinlan, J.R., 2006. Bagging, boosting, and C4.5. *AAAI/IAAI, Vol. 1* (pp. 725-730). RTE News, 2015. Paris attacks death toll rises to 130. [online] Available at: <https://www.rte.ie/news/2015/1120/747897-paris/> [Accessed 24 September 2017].
- Rozin, P. and Royzman, E.B., 2001. Negativity bias, negativity dominance, and contagion. *Personality and social psychology review*, 5(4), pp.296-320.
- Silva, L., Goel, L. and Mousavidin, E., 2009. Exploring the dynamics of blog communities: the case of MetaFilter. *Information Systems Journal*, 19(1), pp.55-81.
- Sherwood, H., 2016. Incidents of anti-Muslim abuse up by 326% in 2015, says Tell MAMA. [online] Available at: <https://www.theguardian.com/society/2016/jun/29/incidents-of-anti-muslim-abuse-up-by-326-in-2015-says-tell-mama> [Accessed 20 September 2017].
- Sky News, 2017. Is Islamophobia on the rise in the UK? [online] Available at: <http://news.sky.com/story/is-islamophobia-on-the-rise-in-the-uk-10921247> [Accessed 20 September 2017].
- Text Mining, Analytics & More., 2014. What are N-Grams?[online] Available at: <http://text-analytics101.rxnlp.com/2014/11/what-are-n-grams.html> [Accessed 20 September 2017].
- Total Politics, 2010. Top 50 UK political blogs. [online] Available at <https://www.tot-politics.com/articles/news/top-50-uk-political-blogs> [Accessed 5 July 2017].
- Vanparys, N., Jacobs, D. and Torrekens, C., 2013. The impact of dramatic events on public debate concerning accommodation of Islam in Europe. *Ethnicities*, 13(2) (pp.209-228.)
- Vehovar, V., Toepoel, V. and Steinmetz, S., 2016. Non-probability Sampling. *The SAGE Handbook of Survey Methodology*, (pp.329-345.)

- Williams, M., 2017. Political Blogs UK Top 10. [online] Available at: <http://www.vu-elio.com/uk/social-media-index/top-10-uk-political-blogs/> [Accessed 6 July 2017].
- Worstell, T., 2013. Explaining The Extraordinary Rise Of The UK Independence Party. [online] Available at: <https://www.forbes.com/sites/timworstell/2013/05/06/explaining-the-extraordinary-rise-of-the-uk-independence-party/#3baa94b85f0f> [Accessed 28 August 2017].