From Brussels to Brexit: Islamophobia, Xenophobia, Racism and Reports of Hateful Incidents on Twitter

Research Prepared for Channel 4 Dispatches - 'Racist Britain'

Carl Miller
Francesca Arcostanzo
Josh Smith
Alex Krasodomski-Jones
Susann Wiedlitzka
Rooham Jamali
Jack Dale

Centre for the Analysis of Social Media, Demos

DEMOS

OVERVIEW

The Centre for the Analysis of Social Media (CASM) at Demos conducted research featured in the Channel 4 Dispatches Documentary Racist Britain**,** first shown on 11th July, 2016. The research focussed on the reaction on Twitter to two significant events: the first was the terrorist attacks on Zaventem Airport and Maalbeek Metro station in Brussels on March 22nd, 2016, and the second was the announcement of the decision of the UK by referendum to leave the European Union on June 24th, 2016.

The overall objective of the research was, against the background of the general reaction on Twitter to these two important events, (a) to understand the amount of Islamophobic, racist, xenophobic or hateful views expressed on Twitter; and (b) understand the role of Twitter as a forum promoting solidarity and support for migrants, and ethnic and religious minorities, and as a platform for those themselves the victims of hate crime. The specific research questions were:

For the EU Referendum:

- to what extent did immigration as a issue feature in the pro-leave and pro-remain campaigning on Twitter in the run-up to the referendum?
- Over the period of the announcement of the referendum result, what was the broad volume of UK-based discussion of migration and migrants on Twitter? To what extent was this discussion explicitly linked to the EU referendum, and how were migrants and migration discussed?
- What was the volume of explicit expressions of xenophobic views on Twitter sent from the UK over the period of the referendum? To what extent did the announcement of a Brexit result coincide with an increase in Islamophobic messages that could be measured?
- How many Islamophobic views were sent on Twitter sent from the UK over the period of the referendum? To what extent did the announcement of a Brexit result coincide with an increase Islamophobic messages that could be measured?
- To what extent was Twitter also used as a forum to show solidarity with and express support for migrant groups, and racial, religious and ethnic minority groups?
- To what extent was Twitter used by the possible victims of xenophobic, Islamophobic or racist abuse to draw attention to their experiences?

For the terrorist attacks in Brussels:

- What was the broad volume of reaction to the Brussels terrorist attacks on Twitter?
- Of the broad reaction, what were the different ways that people used Twitter to react to the Brussels attacks? To what extent did this reaction include discussion of immigration, integration and Islam?
- To what extent did the reaction to the Brussels attacks on Twitter include a discussion linking wider issues related to a perceived lack of integration, assimilation, or issues within Islam itself?
- How many Islamophobic views were sent on Twitter sent from the UK over the period of the Brussels attacks? To what extent did the Brussels attacks coincide with an increase Islamophobic messages that could be measured?
- What was the nature of the Islamophobic messages that could be identified?

RESULTS

For the EU Referendum:

The overall reaction

- Immigration was one of the key themes made by Brexit supporters on Twitter in the run-up to the vote. Between 27th May and 2nd June, 28% of Tweets related to the EU campaign, and sent by Twitter accounts judged to be pro-Brexit, were about immigration.

- Between 22nd June and 30th June, 258,553 Tweets were sent from the UK containing the words 'migrant', 'migrants', 'immigrant', 'immigrants', 'refugee', and 'refugees'. General discussion about migrants and migration sharply increased over the day of the Brexit announcement.

Xenophobia

- Between 19th June and 1st July, 16,151 Tweets were sent from the UK containing a word or hashtag related to xenophobia (see annex). 13,236 of these Tweets were sent between 24th June and 1st July. The day of the Brexit announcement itself saw the highest volume of these Tweets within this period.

- Not all Tweets containing a word related to xenophobia express a xenophobic view, of course. Many people used these words and hashtags to criticise xenophobia. Of the 16,151 Tweets using a xenophobic term sent from the UK, 5,484 were classified as derogatory and xenophobic. 10,671 were classified as not xenophobic (and often angrily denounced xenophobia).[1]

    - On June 24th, 707 Tweets were classified as xenophobic, and 3,549 as supportive.
    - On June 25th, 502 Tweets were classified as xenophobic, and 2,225 as supportive.

Islamophobia

- Between 18 March and 30 June, 2016, 4,123,705 Tweets were sent around the world containing a word that can be used in an Islamophobic way (see annex).[2]

- Of these 4,123,705 Tweets, 28,034 were judged to from the UK and were classified as explicitly anti-Islamic and derogatory.

- A small peak of the volume of derogatory anti-Islamic Tweets occurred between 8pm on June 23rd and midnight on June 25th.
    - 479 derogatory anti-Islamic Tweets were sent on June 24th
    - 146 derogatory anti-Islamic Tweets were sent on June 25th

Reports of offline abuse and hate crime on Twitter

- Between June 25th and July 4th, 98,948 Tweets were sent containing either the hashtag #postrefracism or #safetypin.

- Of these 98,948 Tweets, 81,688 were classified as 'general awareness' - Tweets generally expressing solidarity with migrant groups, or expressing a general concern with a possible increase of hate crime in the post-Brexit period.

- However, the remaining 17,260 Tweets using either #postrefracism or #safetypin were classified to be either (a) providing direct accounts of specific incidents of hate crime or hateful abuse; or (b) relaying other accounts of specific incidents of hate crime or hateful abuse.[3]

---

[1] N.B. Four Tweets were doubly classified as borderline cases.

[2] Most of these words were Islamophobic slurs. It is of course possible to express an Islamphobic view without using an Islamophobic slur; so the numbers presented here are not claimed to be exhaustive of the total amount of xenophobic views expressed on Twitter over this time period.

[3] It should be noted, of course, that none of the accounts of specific incidents of hate crime or hateful abuse could be verified.

- Of these 17,260 Tweets providing either a direct, or relayed, account of specific incidents of hate crime or hateful abuse, 14,847 were Re-tweets, and 2,413 were original messages.
- Of these 17,260 Tweets providing either a direct, or relayed, account of specific incidents of hate crime or hateful abuse, it was judged (see annex) that 5,690 came from another country, 6,720 came from the UK.

For the Brussels Attacks:

The Overall Reaction

- From the start of the Brussels attacks and over the next five weeks, around 13,800,000 Tweets were sent about Brussels and in reaction to the attacks.[4] Around 7,000,000 were sent on the day of the attack itself.

- The 8,452,661 Tweets that contained the words 'Brussels' or 'Bruxelles' were algorithmically classified into the following categories:
  - 32.2% of Tweets expressing solidarity with migrant groups, underlining the strength of common humanity, and rejecting the ability of terrorists to divide communities.
  - 24.1% of Tweets sharing breaking news of the attack itself.
  - 18.2% of Tweets expressing condolences with the victims of the attacks, their family and friends, and the people of Brussels.
  - 8% of Tweets sharing ongoing news coverage of the attacks.
  - 5.8% of Tweets that linked the attacks with the broader issues of immigration, integration and Islam.

(11.7% of Tweets could not be placed into any of these categories).

General Discussion of Integration and Islam

- The 13,800,000 Tweets sent in reaction to Brussels used a very large number of hashtags. Of the 500 most shared, 50 explicitly referred to Islam.

- The most popular hashtag was #stopislam. Across March 22nd and 23rd, 327,391 Tweets used it. However, the hashtag was used to host both messages defending Muslim communities and Islam in the wake of the attacks, and also those critical of them. Of Tweets using the hashtag:
  - 180,000 were classified as being defensive of Muslims and Islam
  - 146,000 were classified as expressing a view that broadly pointed to issues within Islam, migration or a lack of assimilation as underlying contributors to the Brussels attacks.[5]

  - 157,000 Tweets were sent containing a hashtag that was explicitly supportive of Islam.
  - 49,000 Tweets were sent containing a hashtag that was explicitly critical of Islam.

Islamophobia

- From March 22nd to March 30th, 58,074 Tweets were sent containing a word that can be used as an anti-Islamic slur (see Annex).

- Of these 58,074 Tweets containing a word that can be used as an anti-Islamic slur, 4,798 were classified as angry, severely derogatory and explicitly anti-Islamic.
  - Over the two days before the Brussels attacks, an average of 216 derogatory anti-Islamic Tweets were sent per day. In the days after the attacks, an average of 680 derogatory, anti-Islamic Tweets were sent per day.

- A random sample of 100 Tweets classified as angry, severely derogatory and explicitly anti-Islamic were qualitatively analysed, and five categories of Islamophobia were identified:

---

[4] This included data from a number of different collections: (1) Tweets containing the words 'Brussels' or Bruxelles', (2) Tweets sent on a number of related hashtags created in reaction to the attacks, and Tweets containing Islamophobic slurs. (see annex)

[5] N.B. There is no suggestion that these views are necessarily anti-Islamic or derogatory.

- ‘Islam is the Enemy’ - Views that considered Islam as inherently permissive of violence and fundamentally hostile to the West.
- ‘Muslims are Pedophiles’ - Views insinuating that Muslims are more likely to commit sex crimes.
- ‘They’re all terrorists’ - Views suggesting that the wider Muslim populations are permissive and supportive of terrorism.
- ‘Calls for Action - A small number of Tweets seeking to organise offline action against Muslim communities in the wake of the Brussels attacks.
- ‘Anger and Abuse’ - Abusive, derogatory anti-islamic comments, some sent directly to other Twitter accounts (in some cases users intended to be the targets of this abuse).

OVERALL METHODOLOGY

Twitter data is often challenging to analyse. Data drawn from social media are often too large to fully analyse manually, and also often not amenable to the conventional research methods of social science. The research team used a technology platform called Method52, developed by CASM technologists based at the Text Analytics Group at the University of Sussex.[6] It is designed to allow non-technical researchers to analyse very large datasets like Twitter.

Data Collection
Method52 was used to directly collect Tweets from Twitter’s Stream and Search ‘Application Programming Interfaces’ (or APIs). They allow all Tweets to be collected that contain one of a number of specified keywords. The keywords used in the various collections used in this research are detailed in the annex.

Data Analysis
Method52 allows researchers to train algorithms to split apart (‘to classify’) Tweets into categories, according to the meaning of the Tweet, and on the basis of the text they contain. To do this, it uses a technology called natural language processing. Natural language processing is a branch of artificial intelligence research, and combines approaches developed in the fields of computer science, applied mathematics, and linguistics. An analyst ‘marks up’ which category he or she considers a tweet to fall into, and this ‘teaches’ the algorithm to spot patterns in the language use associated with each category chosen. The algorithm looks for statistical correlations between the language used and the categories assigned to determine the extent to which words and bigrams are indicative of the pre-defined categories. Details about how these algorithms were used, and how well they worked, are provided below.[7]

The Accuracy of Algorithms
To measure the accuracy of algorithms into the categories chosen by the analyst, we used a ‘gold standard’ approach. For each, around 100 tweets were randomly selected from the relevant dataset to form a gold standard test set for each classifier. These were manually coded into the categories defined above. These tweets were then removed from the main dataset and so were not used to train the classifier.

As the analyst trained the classifier, the software reported back on how accurate the classifier was at categorising the gold standard, as compared to the analyst’s decisions. On the basis of this comparison, classifier performance statistics – ‘recall’, ‘precision’, and ‘F-score’ are created and appraised by a human analyst. Each measures the ability of the classifier to make the same decisions as a human in a different way:

Overall accuracy: This represents the percentage likelihood of any randomly selected Tweet within the dataset being placed into the appropriate category by the algorithm. It is based on three other measures (below).

---

[6] This group is led by Professor David Weir and Dr Jeremy Reffin. More information is available about their work at: http://users.sussex.ac.uk/~davidw/styled-3/

[7] For a more detailed description of this methodology, see the Demos paper *Vox Digitas,*

Recall: The number of correct selections that the classifier makes as a proportion of the total correct selections it could have made. If there were 10 relevant tweets in a dataset, and a relevancy classifier successfully picks 8 of them, it has a recall score of 80 per cent.

Precision: This is the number of correct selections the classifiers makes as a proportion of all the selections it has made. If a relevancy classifier selects 10 tweets as relevant, and 8 of them actually are indeed relevant, it has a precision score of 80 per cent.

F-Score: All classifiers are a trade-off between recall and precision. Classifiers with a high recall score tend to be less precise, and vice versa. The 'overall' score reconciles precision and recall to create one, overall measurement of performance for each decision branch of the classifier.

N.B. the values for each algorithm (called a classifier) are presented within the detailed methodology of this report. The values are expressed as value up to 1: a value of 0.76, for instance, indicates a 76% accuracy.

CAVEATS

The research of large social media datasets is a reasonably new undertaking. It is important to set out a series of caveats related to the research methodology that the results must be understood in the light of:

- **The algorithms used are not perfect:** throughout the report, some of the data will be mis-classified. The technology used to analyse tweets is inherently probabilistic, and none of the algorithms trained and used to produce the findings for this paper were 100% accurate. The accuracy of all algorithms used in the report are clearly set out in this report.

- **Some data will be missed:** Acquiring Tweets on the basis of the keywords that they contain presents two possible problems. First, the initial dataset may contain tweets that are irrelevant to the thing being studied. Secondly, it may miss tweets that are relevant to the thing being studied. Researchers worked to construct as comprehensive a list of keywords as possible (these are detailed in the report, below), however it is likely some were missed, and the numbers presented in this report are likely a subset of the total.

- **Twitter is not a representative window into British society**: Twitter is not evenly used by all parts of British society. It tends to be used by groups that are younger, more socio-economically privileged and more urban. Additionally, the poorest, most marginalised and most vulnerable groups of society are least represented on Twitter; an issue especially important when studying the prevalence of xenophobia, Islamophobia and the reporting of hate incidents.[8]

- Overall, this research is intended to be an **indicative**, first-take of the reaction on Twitter to these important events. It is not presented as either exhaustive or definitive; and it is very much hoped that it will stimulate further research on this vital topics in the future.

[8] For a longer discussion of this issue, see the Demos paper *The Road to Representivity*

DETAILED METHODOLOGY

The EU Referendum

This section details the method used to research Twitter over the period of the announcement of the EU referendum result.

PART I: The Referendum Campaign

From May 20th to June 2nd we collected all tweets sent to British MPs containing a keyword or a hashtag related to the EU Referendum (see the appendix). As a first step, we have created a campaign classifier in order to distinguish between pro-leave and pro-remain tweets.

Category	Precision	Recall	F-Score	Accuracy
Pro-Leave	0.857	0.757	0.804	0.693
Pro-Remain	0.629	0.629	0.629	0.693
Other	0.351	0.520	0.419	0.693

Tab. 1. Reliability scores for classifier separating pro-leave and pro-remain Tweets

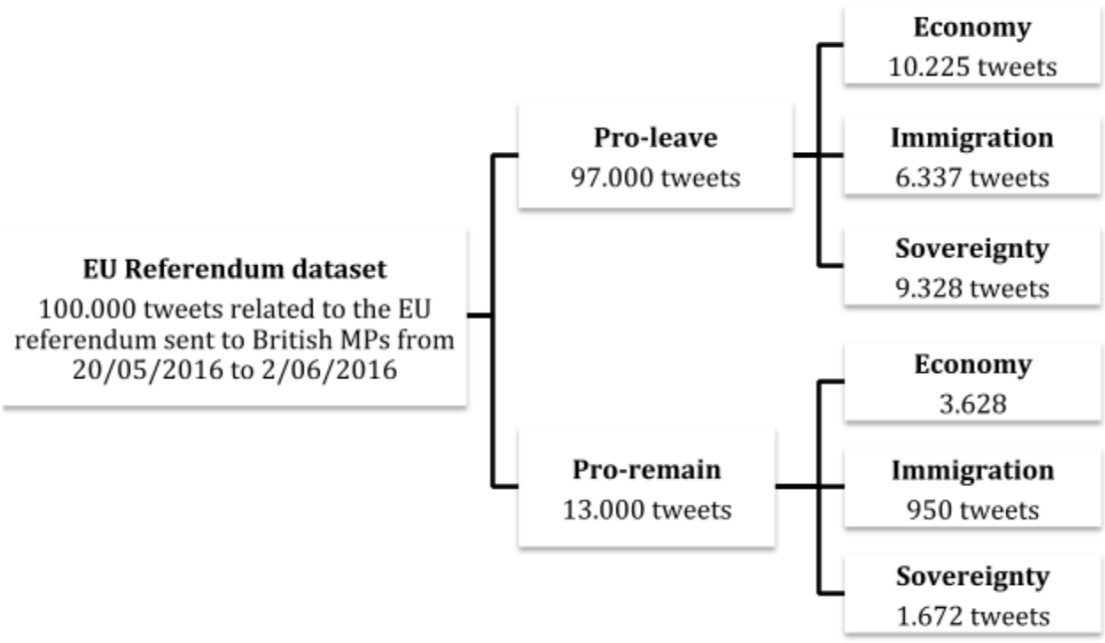
(hence, the overall likelihood of any given Tweet being correctly categories was 69.3%. The algorithm was better at classifying pro-leave and pro-remain tweets; less good at classifying 'other' tweets).

As a second step, we have created an issue classifier in order to distinguish, for both pro-Brexit and pro-Remain Tweeters, which were the most relevant issues into the debate.

Category	Precision	Recall	F-Score	Accuracy
Sovereignty	0.810	0.586	0.680	0.775
Economy	0.793	0.561	0.657	0.775
Immigration	0.286	0.667	0.400	0.775
Other	0.790	0.890	0.837	0.775

Tab. 2. Reliability scores for classifier separating EU referendum campaign Tweets by the issues they primarily mentioned

Together with economy and sovereignty, immigration resulted to be one of the three issues mostly associated with the referendum debate both for pro-Remain and pro-Brexit Tweeters.



Last, all Tweets were collected sent from the UK that contained one of the following words: 'migrant', 'migrants', immigrant', immigrants', 'refugee' or refugees'. This resulted in a collection of 258,553 tweets. From this dataset, we have been able to identify a sub-sample of 40,255 tweets containing a keyword linked to Brexit (see the annex for a list of these keywords).

PART II: Xenophobia on Twitter

In order to measure the number of Tweets containing xenophobic views from the UK over the announcement of the Brexit result, first all Tweets were collected between 19th June and 1st July that contained a term, including hashtags, that could be related to xenophobia. This resulted in a dataset of 16,151 tweets based in the UK. See the annex for a list of these keywords. N.B. this is not comprehensive, and may be only a small amount of the total number of xenophobic Tweets that were sent over this period.

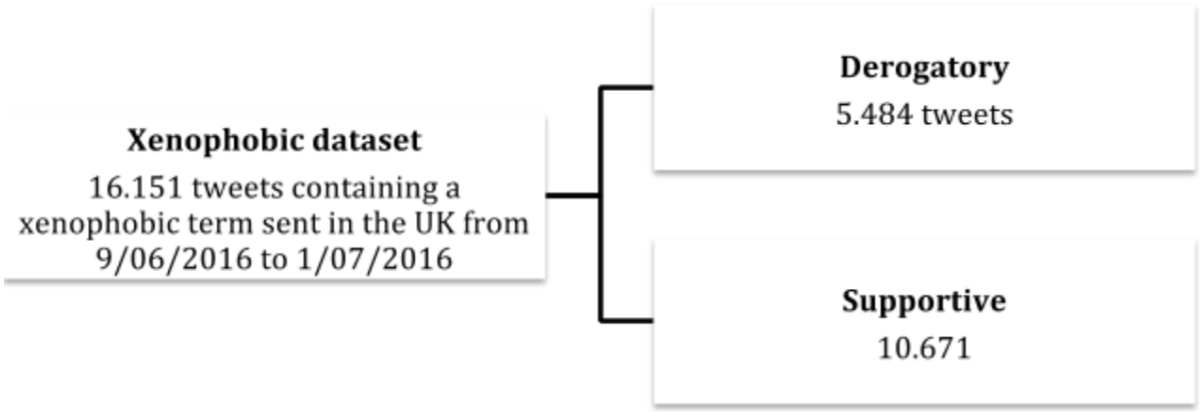
However, not all Tweets that contain a xenophobic term express a xenophobic attitude. Xenophobic words and hashtags can be used in a broad number of different ways, including to dispute xenophobic views and express support for migrants. In order to measure the extent of Tweets expressing xenophobic views, a 'xenophobia classifier' was trained to distinguish between tweets that expressed derogatory and xenophobic views towards migrants, and those that didn't.

Category	Precision	Recall	F-Score	Accuracy
Derogatory	0.806	0.941	0.868	0.760
Supportive	0.222	0.133	0.167	0.760
Other	0.000	0.000	0.000	0.760

Tab. 3. Reliability scores for classifier dividing Tweets containing possibly xenophobic terms into those expressed a derogatory view towards migrants and those that expressed a view supportive of migrants

(The classifier was better at identifying derogatory Tweets than supportive ones. The 'other' category was not used for classification The capacity of the classifier to identify derogatory Tweets was its most important function. Overall, any given Tweet had a 76% chance of being correctly classified)

The outcome of the classifier showed that, of Tweets using xenophobic terms, roughly twice as many Tweets were supportive of migrants and disputing xenophobic attitudes as were xenophobic.



PART III: Islamophobia on Twitter

The identification of Islamophobic Tweets in the wake of the Brussels attacks formed part of a longer-term research process. It is detailed within this report, below, within the methodological description of the identification of Islamophobic Tweets in the wake of the Brussels attacks.

PART IV: Reports of hateful incidents

In the aftermath of the EU referendum, we collected all tweets containing the hashtags #SafetypPin and/or #PostRefRacism. This resulted in a dataset of 98.948 tweets sent from 25th June to July 4th. The function of these hashtags were to show solidarity with migrants in the wake of the brexit decision, and also to raise awareness of hateful incidents (whether xenophobic, Islamophobic, or racist) taking place.

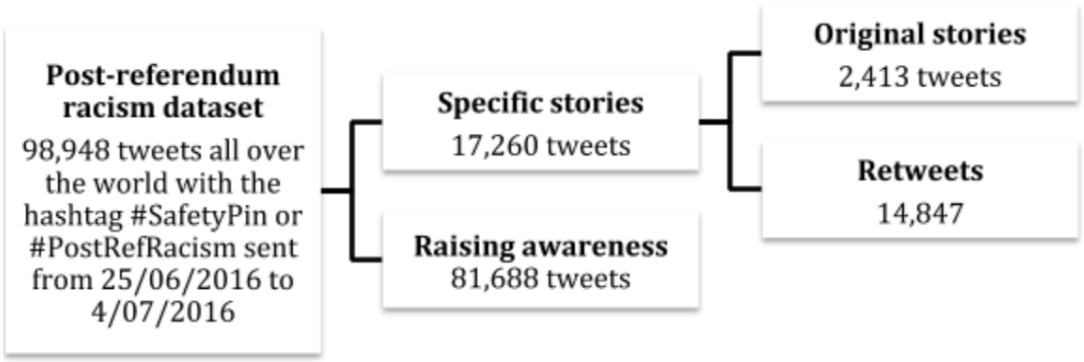
In order to understand the number of accounts of specific and hateful incidents, a classifier was trained to distinguish between tweets generally raising awareness about hate crime, from tweets containing specific accounts of hateful incidents.

Category	Precision	Recall	F-Score	Accuracy
Specific stories	0.571	0.421	0.485	0.830
Raising awareness	0.872	0.926	0.898	0.830

Tab. 4. Reliability scores for separating Tweets sharing specific accounts of hateful incidents with those generally raising awareness about an increase in hate crime

(the 'recall' of the algorithm for specific stories is reasonably low. This was accepted as reasonable because it meant that the classifier was unlikely to exaggerate the number of specific stories identified).

The algorithm classified 17,260 Tweets as containing an account of a specific incident, and 81,688 as generally raising awareness of hate crime and hateful incidents. Of these 17,260 tweets that contained an account of a specific incident, 2413 were original tweets, and the rest - 14,847 - were retweets of these original tweets.



N.B. Upon qualitative evaluation, the Tweets within the specific stories dataset included both first-hand accounts of hateful incidents, but also accounts second-hand, or that the Tweeter had heard about from another source.

The Brussels Attacks

This section details the method used to research Twitter over the period of the announcement of the Brussels attacks.

PART I - The Wider Reaction

To understand the overall reaction to the Brussels attacks, all Tweets were collected that contained the words 'Brussels' or 'Bruxelles' sent between 14 March and 31 March 2016. This returned 8,452,661 Tweets. This large dataset was analysed using a series of algorithms.

First, a classifier was trained to identify Tweets simply sharing news concerning the attacks. It split Tweets into three categories:

- 'News': those sharing news the attacks had taken place;
- 'Condolence': those expressing condolence for those affected by them;
- 'Other': all other Tweets.

Category	Precision	Recall	F-Score	Accuracy
News	0.690	0.967	0.804	0.780
Condolence	0.762	0.842	0.800	0.780
Other	0.892	0.647	0.750	0.780

Tab. 5. Reliability scores for classifier separating Tweets sharing news about the attacks and expressing condolences to those affected by them with all other Tweets

Second, a classifier was trained to sort all Tweets identified as 'Other' (see above) into the following categories:

- 'Solidarity': Tweets that called for religious tolerance, that expressed solidarity with Muslim communities, and expressed the view that the terrorist attacks were in no way linked to Islam or migration.
- 'Criticism and wider issues': Tweets that linked the terrorist attacks to wider issues within Islam, migration or integration, that called for Muslims to be more active in combatting terrorism, and that expressed support for Donald Trump's views on halting immigration of Muslims in the wake of the attacks.
- News: Tweets covering new details of the attacks and attackers as they became known.
- 'Rest': Tweets that did not fall into any of the other categories

Category	Precision	Recall	F-Score	Accuracy
Solidarity	0.667	0.235	0.348	0.510
Criticism and wider issues	0.583	0.718	0.615	0.510
News	0.714	0.476	0.571	0.510
Rest	0.321	0.391	0.353	0.510

Tab. 6. Reliability scores for classifier separating Tweets on the basis of their broad nature of reaction to the Brussels attacks

PART II - #StopIslam

In order to analyse the important hashtag #stopislam, that was used before the Brussels attacks, but because important in the reaction to them, all Tweets were collected that contained '#stopislam' between 14th March and 31st of March. This returned 697,836 Tweets. 420,635 of these Tweets were written in English.

A classifier was trained to separate all the English-language Tweets using #stopislam in this dataset into those that were:

- 'Critical' of Islam: that pointed to wider aspects of Islam or Muslim communities, their faith or culture, as being in some way implicated in the attacks. N.B. by no means were all these views derogatory, hateful or anti-Islamic.
- 'Supportive' of Islam: that defended Islam and Muslims, and disputed the critical views expressed in Tweets using the hashtag.
- 'Other ' - Tweets that did not fit within either of the two categories, above.

Category	Precision	Recall	F-Score	Accuracy
Critical	0.794	0.810	0.802	0.754
Supportive	0.711	0.831	0.766	0.754
Other	0.6677	0.111	0.190	0.754

Tab. 7. Classifier separating Tweets sent on #stopIslam into those broadly critical of Islam, and those broadly supportive of Islam

Part 3 - Islamophobia

Identifying Tweets that were hateful, derogatory and anti-Islamic was a formidable analytical challenge.

First, all Tweets were collected that contained one of an extensive list of terms that could be used in an anti-Islamic way (see annex). This collection began on the 18th March and continued until the 30th of June. It returned a very large number of Tweets over this period: 4,123,705. The very large majority of these Tweets were not anti-Islamic or hateful. A series of algorithms were built to respond to the different challenges that this dataset posed. Each was designed to remove Tweets which were not Islamophobic from the dataset:

- A large number of Tweets contained the word 'Paki'.[9] A classifier was used to separate derogatory uses of this word from non-derogatory uses.
- A large number of Tweets also contained the word 'terrorist'. Of course, many Tweets containing this word were in now way derogatory or anti-Islamic. Two classifiers were built to analyse tweets containing these words:
  - First, a classifier was trained to separate Tweets referring to Islamist terrorism from other forms of terrorism.
  - Second, of the Tweets referring to Islamist terrorism, a classifier to distinguish views broadly attacking Muslim communities in the context of terrorism, from those broadly defending Muslim communities.

[9] N.B. whilst this word refers to an ethnic rather than religious group, it was found that it was often used interchangeably to refer to Muslim communities

- A classifier was trained to separate all other Tweets in the dataset into those that were derogatory and anti-Islamic from those which were not.
- Last, the Tweets that, based on the above, (a) used the term 'Paki' in a derogatory way, (b) that used the term 'terrorist' to broadly attack Muslims or Muslim communities, (c) that used the other possible slur terms in the collection in a way that was anti-Islamic were combined. These were then filtered to include only Tweets sent from the UK. This resulted in the final total of Islamophobic Tweets.

The accuracy of these algorithms are as follows:[10]

Category	Precision	Recall	F-Score	Accuracy
Hate	0.680	0.745	0.711	0.715
Other	0.753	0.689	0.719	0.715

Tab. 9. Reliability scores for classifier identfiying Tweets using 'Paki' in a derogatory way

Category	Precision	Recall	F-Score	Accuracy
Islamist Terrorism	0.870	0.762	0.812	0.858
Non-Islamist Terrorism	0.852	0.923	0.886	0.858

Tab. 10. Reliability scores for classifier separating references to Islamist and non-Islamist terrorism

Category	Precision	Recall	F-Score	Accuracy
Attacking	0.645	0.727	0.648	0.693
Defending	0.718	0.830	0.770	0.693
Other	0.800	0.291	0.427	0.693

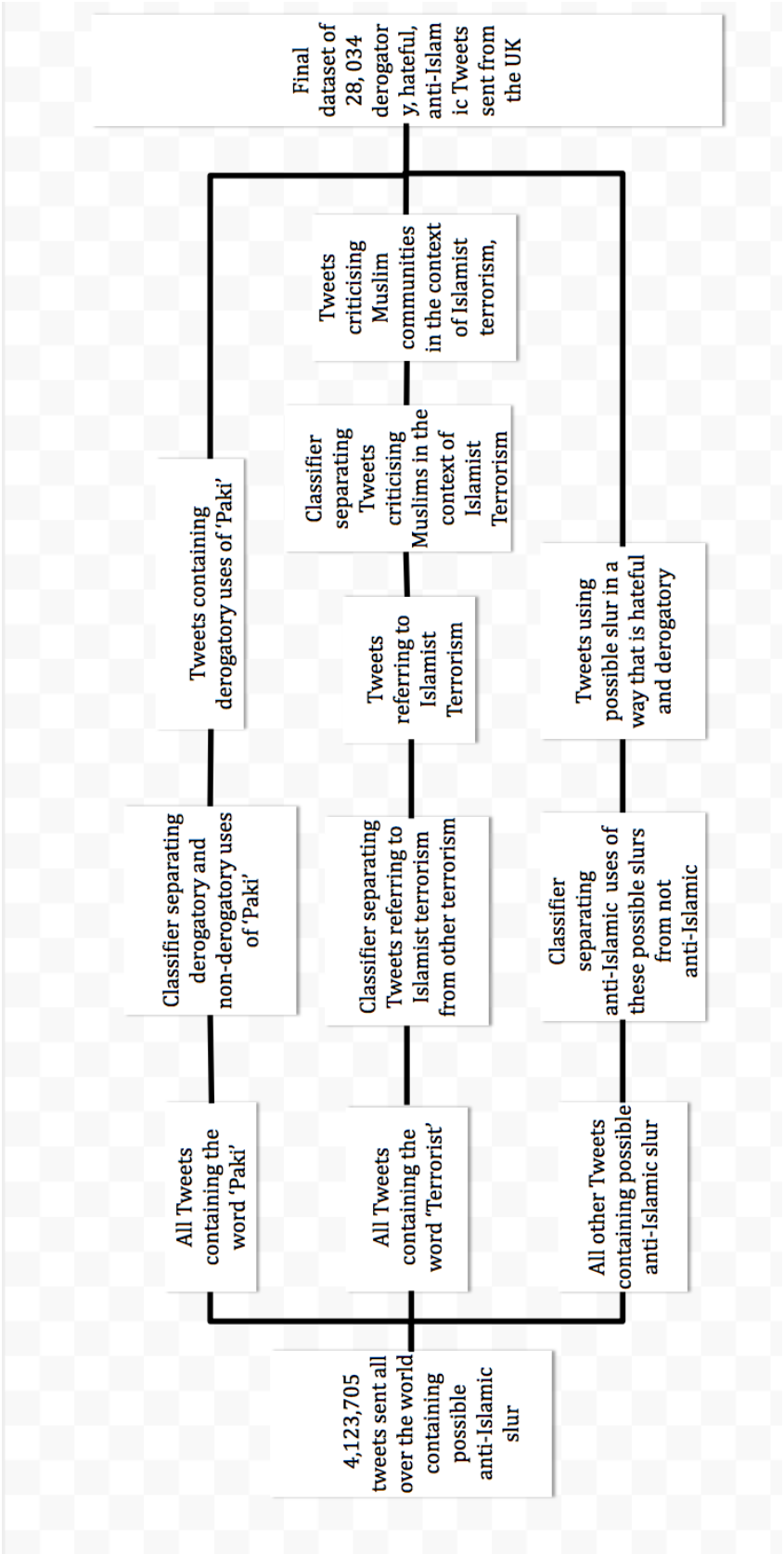
Tab. 11. Reliability scores for classifier separating Tweets attacking or defending Muslims in the context of a reference to terrorism

Category	Precision	Recall	F-Score	Accuracy
Anti-Islamic	0.533	0.702	0.606	0.740
Not anti-Islamic	0.864	0.755	0.806	0.740

Tab. 12. Reliability scores for classifier separating Tweets containing other possible slurs into those that were hateful and anti-Islamic from those that were not

[10] Due to the large number of classifiers used, the accuracy was checked by taking a random sample of Tweets that - according to the system of algorithms described above - were classified as derogatory anti-Islamic. 78 of these 100 were identified by an analyst as derogatory and anti-Islamic.

These algorithms were connected together into an 'architecture', shown below. Each Tweet collected passed through the architecture on the basis of how it was classified. Overall, this system of algorithms succeeded in filtering the very large (4.12 million) number of Tweets into a much smaller (28,034) subset that were much more likely to be hateful, derogatory and anti-Islamic.



Annex - Data Collection Keywords

The annex contains the keywords used to collect Tweets analysed throughout this report.

I. The referendum campaign. List of used hashtags related to the EU referendum:

- #euref
- #eureferendum
- #brexit
- #voteleave
- #strongerin
- #strongerout
- #voteremain
- #votestay
- #takecontrol
- #remain
- #Leave

II. UK-based general discussion of migration

- migrant
- migrants
- immigrant
- immigrants
- refugee
- refugees

III. Words/Hashtags used to identify Tweets within the general discussion of migration (collected using terms described above, in Annex II) explicitly linked to the EU Referendum

- brexit
- European Union
- EU
- Ukip
- @UKIP
- #strongerin
- #remain
- #voteremain
- #votein
- #bremain
- #intogether
- #voteleave
- #leaveeu
- #takecontrol
- #go
- #leave
- #betteroffout
- Farage
- Boris Johnson
- @Nigel_Farage
- @BorisJohnson
- #safetypin
- #PostRefRacism
- #PolesinUK
- #londonstays

- #EUref

IV. Words/Hashtags used to collect Tweets that could be related to Xenophobia

- #refugeesnotwelcome
- #defendEurope
- #Whitegenocide
- #whitepower
- #whitepride
- Illegals
- #whiteresistance
- #whiterevolution
- #sendthemhome
- Refugees not welcome
- #MakeBritainwhiteagain
- #sendthemback
- #Getoutwevotedleave
- #Stopimmigration
- #DeportallMuslims
- #NeverIslam
- Rapefugee
- #Polesgohome
- #NojobsinUKforEU
- golliwog
- Rapeugee
- #fuckislam
- #PolishVermin
- #NoIslam
- #BanIslam
- muzzrats
- muzzle
- Londonistan
- muscat
- muzrat
- muzzrat
- musrats
- immigrants go home
- refugees go home
- migrants go home
- curry munching
- dirty pack
- anti-immigrant
- anti-immigration
- #NoMoreRefugees
- #StopTheInvasion
- #NoMoreMigrants
- #EndIslam
- #IslamIsTheProblem
- Rag head

V. Words/Hashtags used to collect Tweets that could be derogatory and anti-Islamic

- Jihad
- Jihadi
- Sand Flea
- Terrorist
- hijab
- Camel Fucker

- Carpet Pilot
- Clitless
- Derka Derka
- Diaper-Head
- Diaper Head
- Dune Coon
- Dune Nigger
- Durka-durka
- Jig-Abdul
- Muzzie
- Q-Tip Head
- Rab
- Racoon
- Rag-head
- Rug Pilot
- Rug-Rider
- Sand Monkey
- Sand Moolie
- Sand Nigger
- Sand Rat
- Slurpee Nigger
- Towel-head
- Muslim Paedos
- Muslim pigs
- Muslim scum
- Muslim terrorists
- Muzrats
- muzzies
- Paki
- Pakis
- Pisslam
- raghead
- ragheads
- Towel head
- FuckMuslims
- WhiteGenocide
- Pegida
- EDL
- BNP
- Rapefugee
- Rapeugee
- mudshark
- kuffar
- kaffir

VI. Search terms for post-referendum hashtags used to report possible hateful incidents

- #SafetyPin
- #PostRefRacism