UPPSALA
UNIVERSITET

IT 15056

Examensarbete 30 hp
Juli 2015

Using machine learning to identify jihadist messages on Twitter

Enghin Omer

Institutionen för informationsteknologi
Department of Information Technology

UPPSALA
UNIVERSITET

Teknisk- naturvetenskaplig fakultet
UTH-enheten

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
http://www.teknat.uu.se/student

Abstract

Using machine learning to identify jihadist messages on Twitter

Enghin Omer

Jihadist groups like ISIS are spreading online propaganda using various forms of social media such as Twitter and YouTube. One of the most common approaches to stop these groups is to suspend accounts that spread propaganda when they are discovered. However, this approach requires that human analysts manually read and analyze an enormous amount of information on social media. In this work we make a first attempt to automatically detect radical content that is released by jihadist groups on Twitter. We use a machine learning approach that classifies a tweet as radical or non-radical and our results indicate that an automated approach to aid analysts in their work with detecting radical content on social media is a promising way forward.

Handledare: Lisa Kaati
Ämnesgranskare: Michael Ashcroft
Examinator: Roland Bol
IT 15056
Tryckt av: Reprocentralen ITC

This thesis is dedicated to all people that are or have been affected by terrorist activities.

Acknowledgement

I want to express my sincere thanks to my thesis supervisor, Dr. Lisa Kaati, for her expert guidance and continuos support and feedback throughout the research. Her guidance helped me to finish this project and report. Thanks to her we wrote a paper on this topic as well. I am also grateful to my reviewer, Dr. Michael Ashcroft, for his feedback and very valuable lessons he gave about machine learning every Friday. Thanks to my co-workers, Amendra Shrestha and Maxime Meyer, for their support and cooperation. Finally, many thanks to Uppsala University and all the people that has helped me to finish this thesis and my master degree.

Contents

List of Figures

List of Tables

1 Introduction

Since the late 1980s, the Internet has become a central and dynamic means for communication, more and more people are using the benefits of Internet worldwide. A wide range of sophisticated technologies has been developed that are connecting people. In 2014 there were over three billion Internet users and the number is still growing [19]. Internet technology comes with numerous benefits including sharing information and ideas as well as accessing them fast and easy. This has created a medium for businesses, consumers, organizations and governments to communicate with each other. It also created a perfect place for various terrorist organizations to disseminate information that aid their causes. There are many different active terrorist organizations in the world today and almost every day newspapers report about terrorist attacks in different parts of the world. According to FBI [18], terrorism has following characteristics:

- involve violent acts or acts dangerous to human life that violate federal or state law
- appear to be intended to intimidate or coerce a civilian population;
- to influence the policy of a government by intimidation or coercion
- to affect the conduct of a government by mass destruction, assassination, or kidnapping

Many terrorist groups use the Internet to spread propaganda. Propaganda usually includes virtual messages, presentations, audio and video files that contain explanations, justifications and/or promotion of terrorist activities. The aim of the propaganda is recruitment and to influence opinion, emotions and attitudes. The availability of terrorist related material on the Internet plays an important role in radicalization processes. Such processes often accompanies the transformation of recruits into individuals determined to act with violence based on extremist ideologies [22]. Another objective of terrorist propaganda is to generate anxiety, fear and panic in a population by releasing violent videos like killing people who fight against terrorist organizations.

One of the most common approaches to stop these groups is to suspend accounts that spread propaganda when they are discovered. However, this approach requires that human analysts manually read and analyze an enormous amount of information on social media.

Detecting radical content in order to react on it or to work with partners to remove it is an important task for law enforcement agencies. The automatic detection proposed in this work should be seen as a complementary way to detect radical content and present it to an analyst for further actions. In this thesis we are addressing the problem of classifying tweets as supporting ISIS or not. We sometimes refer to this as classifying tweets into radical or non-radical even though the problem that we are considering cannot be generalized into solving the problem of detecting radical content in general. This work can be seen as piece in the puzzle of reducing terrorist related material available online. By detecting radical content, such messages can be removed and less people will be exposed to the content. Another use of this work may be to help analysts to detect twitter users that promote radical views.

This report is outlined as follows. Chapter 2 describes how jihadists use social media and provides an introduction to machine learning techniques that are used in this thesis. Chapter 3 presents related work that has been done in the area. Chapter 4 covers details about how the classifier was built including information about features and feature vectors. In chapter 5 the experiments that have been conducted and the results are presented. The work is concluded in Chapter 6 and in Chapter 7 some directions for future work are presented.

2 Theory

2.1 Social media

Social media is a group of Internet-based applications that enable users to create and share content or to participate in social networking [48]. There are many different forms of social media: discussion boards, blogs, microblogs and different kind of networking platforms such as Facebook and Weibo. Twitter is one of the most well-known microblog [30]. Twitter enables users to send and read 140-characters messages called "tweets". To mark different themes and topics in a message it is common to use hashtags. A hashtag is a word or an unspaced phrase prefixed with the hash character (#). This is done to increase the visibility of the tweet. Sometimes in order to promote a product, an idea or a political view, hashtag campaigns are organized which means that hashtags related to a specific topic are intensively used.

2.2 Extremist groups and the use social media

Social media is not only used to communicate with friends and family but also to promote radical views. In many cases individuals and organizations use social media to attract fighters and fundraisers to specific causes. Jihadists, people participating in a jihad [40], have aggressively expanded their use of Twitter as well as other social media applications such as YouTube and Facebook. In 2015 around 90000 Twitter accounts are suspected to support extremist groups [20].

Nowadays it is not necessary to go on the battlefield to join extremist groups and fight with them. Sitting in front of a computer and promoting radical views on social media can be a valuable contribution to promote extremist groups. One example of this is a media mujahideen. Mujahideen is the plural form of mujahid and is used to describe someone involved in jihad [10]. A media mujahideen is formed by people who are fighting on media platforms promoting their extremist propaganda [29].

Since 2011, members of jihadist forums have issued media strategies that encourage the development of a media mujahideen. Guides describing how to use social media platforms and lists of recommended accounts to follow are released in various forums [23]. One such guide is a Twitter guide entitled "The Twitter Guide: The Most Important Jihadi Sites and Support for Jihad and the Mujahideen on Twitter". This guide outlines reasons for using Twitter and states that Twitter is an important arena of the electronic front. The guide has identified 66 important jihadist accounts that users should follow.

One group that is officially recognized as a terrorist group by the United Nations [21] including countries like United States [5], Canada [2], United Kingdom [12] is ISIS. Based on this considerations we assumed that messages posted by users clustering with known ISIS sympathizers contain radical content. In this thesis messages posted on Twitter and promoting ISIS propaganda were used as well as random tweets and tweets having messages against ISIS. This messages were in English and tweets with messages supporting ISIS were collected between 25th of June and 29th of August 2014. The data was used only for research purposes.

ISIS organizes hashtags campaigns showing support for ISIS and its cause. Examples of hashtags are #ILOVEISIS, #ALLEYESONISIS, #ISLAMICSTATE. Another strategy to spread propaganda on Twitter is by using "trending" hashtags even though they are not related to a specific activity (like hashtags related to WorldCup 2014 or IPhone 6) [7].

Organizations active on network like those spreading propaganda on social media are often diffuse, leaderless, and incredibly resilient. That makes tackling terrorist propaganda a difficult task since jihadists have the ability to reorganize themselves all the time. ISIS, for example, uses dispersed forms of network organization and strategies to disseminate rich audiovisual content from the battlefield in near-real time [8]. This fact makes ISIS a challenge for traditionally hierarchical organizations to counter. ISIS has successfully used social media to recruit new members from all over the world [6]. Studying social media seems to be an important approach to identify and understand radical messages.

2.3 Machine Learning

Machine learning is the science that explores how algorithms can be constructed so that they can learn from data and make future predictions [24]. These algorithms build a mathematical model based on some example inputs and use the model to make some predictions or decisions. Using the model any input can be mapped to a range of outputs. The flow of creating the model from the input data and the processes of mapping new inputs to expected outputs is illustrated in Figure 1.

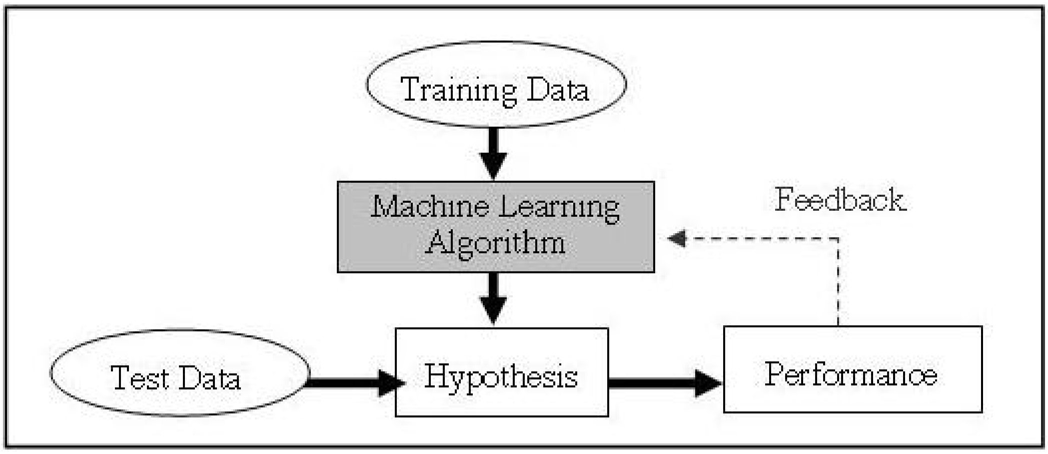

Figure 1: Machine learning algorithm work-flow [9]

The goal is to train models that learn without human intervention or assistance. Instead of static programming that tells the computer what to do, machine learning will construct algorithms that can learn from data. It means that the computer will come up with its own model based on the data provided.

In machine learning, learning methods are classified into three categories:

- Supervised learning
- Unsupervised learning
- Reinforcement learning

The results of experiments are visualized using what is called a confusion matrix. A confusion matrix, known also as a contingency table or error matrix is used in machine learning to visualize the results of a supervised learning algorithm [42]. It consists of two rows and two columns which contain the number of false positives, false negatives, true positives and true negatives instances. The structure of the confusion matrix can be seen in Figure 2.

	p' (Predicted)	n' (Predicted)
p (Actual)	True Positive	False Negative
n (Actual)	False Positive	True Negative

Figure 2: The structure of confusion matrix [1]

The columns of the table represents the instances that the model predicted while the rows represent the instances in the actual class. Based on the confusion matrix a series of more detailed analysis can be done.

- Accuracy is the proportion of the sum of true results and the total number of instances. It shows the percentage of total instances that were correctly classified.

  Accuracy = (true positive + true negative)/(true positive + true negative+false positive + false negative)

- Precision, also called positive predictive value, is the proportion of true positive values within the positive class.

  Precision = true positive/(true positive + false positive)

- Recall is the proportion of positives classified as such.

  Recall = true positive/(true positive+false negative)

2.3.1 Algorithms

SVM (support vector machine) is a large margin classifier. It means that its goal is to find a boundary between two classes that maximizes the distance between the data that are part of this classes as in Figure 3. Depending on how the data is distributed there are different approaches on SVM [49].

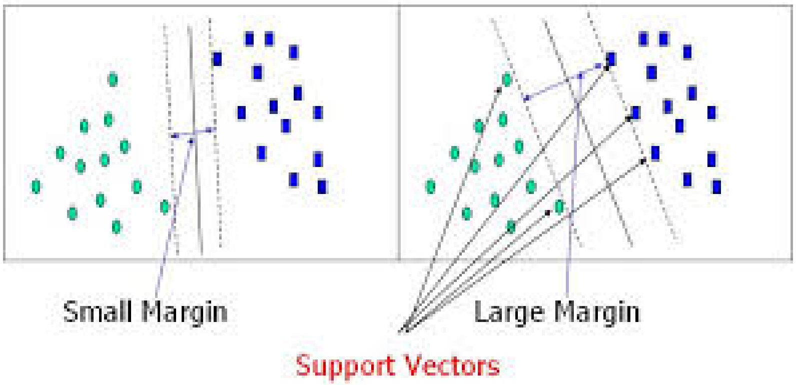

Figure 3: The margin and support vectors [15]

Linearly separable data case involves the fact that there exists at least one hyperplane that can separate the points of the two classes as in Figure 4. The goal of SVM is to find the hyperplane that gives the largest distance to the training examples. The distance is called margin and the optimal hyperplane is the one that maximizes the margin. The points that are closest to the hyperplane are called support vectors.

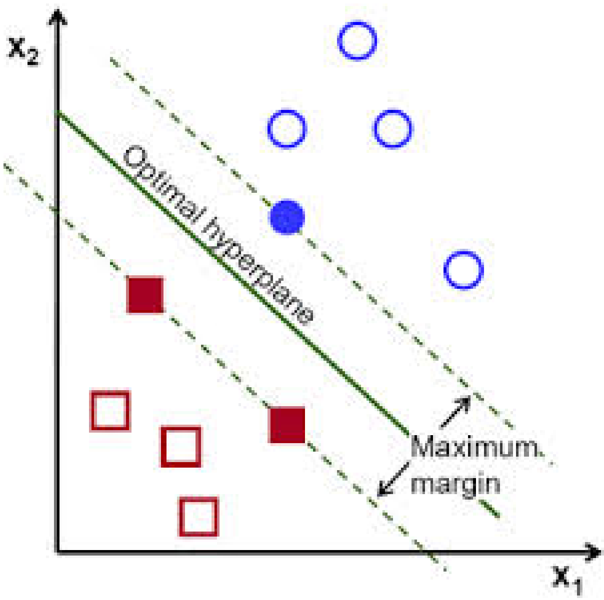

Figure 4: The separable case [17]

The linearly separable case is valid for linearly separable data. Such cases are rare in practice. Often, classes contains points that overlap, so there does not exist a hyperplane that can separate all the classes' points. This means that there will be some samples misclassified. In this case the model needs to include the requirement from the previous case, finding the hyperplane that maximizes the margin, and a new one that minimizes the misclassification errors. A new parameter ε_i is defined for each sample of the training data. It represents the distance of the corresponding training sample to the correct decision boundary as shown in Figure 5. If the sample is in the correct part of the hyperplane then the value of ε_i is 0.

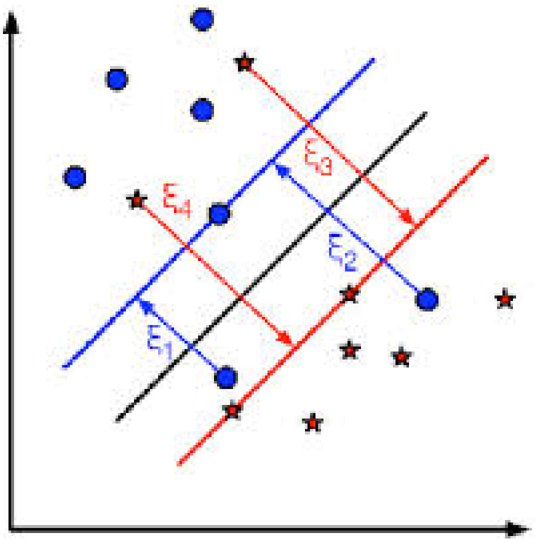

Figure 5: The non-separable case [16]

Another way of solving a non-linearly separable data problem is to map the data to a higher dimensional space and then use a linear classifier as in Figure 6. The way of doing this mapping is called "the kernel trick".

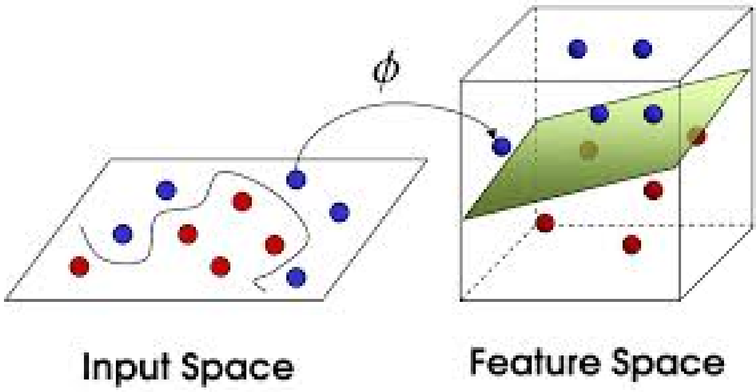

Figure 6: The kernel trick [14]

AdaBoost is a machine learning algorithm that is based on boosting. Boosting is a method that combines moderately inaccurate rules of thumb to create a very accurate classifier. It is based on the assumption that it is easier to find many rules of thumb than a single very accurate one [26]. First an algorithm is defined to find the rules of thumb. A rule of thumb is called a weak learner. The boosting algorithm calls the weak learner repeatedly. Each call generates a weak classifier that separates the inputs like in Figure 7. Decision trees are the most popular weak classifiers used in boosting schemes.

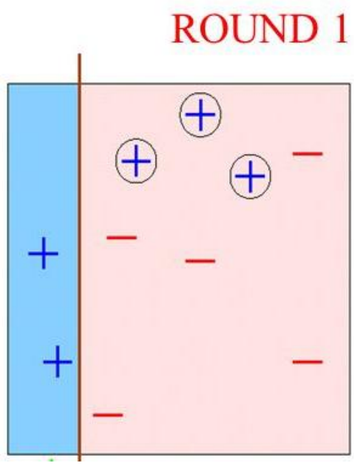

Figure 7: The classification after a weak learner was called [4]

After each call the decision tree algorithm adds a node to its classification tree. The node represents a binary test made on the attributes, and the leafs the label of the data after the decision.

Naive Bayes is a probabilistic classifier based on applying Bayes theorem assuming independence between the features [34]. Naive Bayes computes the probability p of a document d being of class c: $p(c|d)$. Given a document d to be classified, represented by a vector $d = (d_1, ..., d_n)$ the conditional probability can be written using Bayes' theorem as:

$$p(c|d) = \frac{p(c)\, p(d|c)}{p(d)}$$

2.3.2 Text classification

Text classification is the task of classifying a document into a predefined category [31]. In our case the document is a tweet and the category is a Boolean value indicating if the tweet contains radical content or not. As in every supervised machine learning task a dataset is needed. The text classification process includes the following steps:

- read the documents
- preprocess the text (this may include tokenizing the text, lemmatizing, deleting stop words)
- create feature vectors
- select features
- create a model

2.3.3 Dataset

The first step in creating a model is to ensure that a proper dataset is collected. In a raw format a dataset can consist of documents, images, sound recording etc. The data used to construct a model is called the training data. A training sample is a collection of instances $\{x_i\}_{i=1}^n = \{x_1, x_2, ..., x_n\}$

which acts as the input to the learning algorithm for a statistical model where each instance $\{x_i\}$ represents a specific object [50].

In order to examine the performance of the model a test dataset is used which has the same characteristics as the training dataset, that is to say the same features.

2.3.4 Cleaning the data

When data is collected it is usually not "clean" due to several reasons:

- noisy- containing errors or outliers
- misspelled words
- unwanted elements: quotes (retweets), strange symbols

Before using the data it needs to be cleaned. This involves dealing with the missing values, identifying noisy data and correct inconsistent data. Basic methods for dealing with missing items include discarding rows with missing items or estimating the missing item using simple statistical methods such as the mean or median value of the variable whose item is missing [38].

After cleaning the data preprocessing needs to be done. Depending on the situation this may include:

- Stemming (bringing a word to its base form)
- Removing stop words
- Splitting sentences into tokens

2.3.5 Feature vectors

In machine learning a *feature vector* is a n-dimensional vector of numerical features that represents some object [47]. Usually, algorithms that are used in machine learning requires a numerical representation of feature vectors because this facilitates mathematical computation and statistical analysis. The instance is often represented by a n-dimensional feature vector $x = (x_1, ..., x_n) \epsilon R^n$, where each dimension is called a feature. The length n of the feature vector is known as the dimensionality of the feature vector [50].

Examples of features are:

- count of words
- presence of words
- presence of punctuation marks
- count of punctuation marks
- time-based features like the hour when a post was published

2.3.6 Feature selection

Before creating feature vectors a decision on which features should be used needs to be done. Ideally is to use as less features as possible that maximizes the information the model gains. The reasons for using as few features as possible are:

- More features lead to more noise which means irrelevant data is used to create the model.
- There is the risk of *overfitting*. Overfitting occurs when all the data is tried to fit into the model [47]. An example of overfitting can be seen in Figure 8.
- Computational constraints. More features and parameters used in the model will lead to more complex computational operations.

There are several feature selections methods. In this work we have used the *information gain* algorithm to analyze feature selections. Information gain tells us how important a given attribute of the feature vectors is [32]. The way information gain IG is computed for a set of training examples T and an attribute a is as follow:

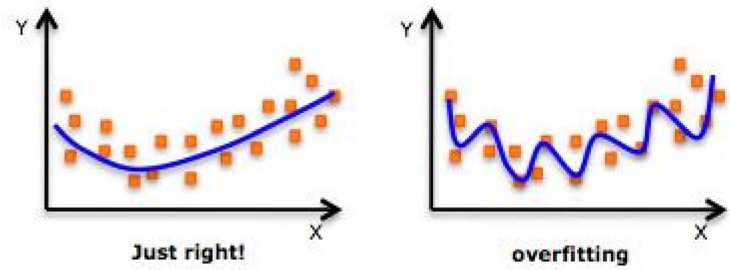

Figure 8: Example of overfitting [11]

$$IG(T, a) = H(T) - H(T \mid a)$$

where T is a set of training examples of form $(x, y) = (x_1, ..., x_n, y)$, x_i is the value of the ith attribute of example x and y is the corresponding class label. $H(T)$ is the information entropy and is computed as follow:

$$-\sum_{i}^{n} P(x_i) \log_b P(x_i)$$

where X is a discrete random variable and $P(X)$ is its probability.

3 Related Work

A lot of research has been done on tweets classification where tweets are classified into several classes as in [25] where tweets are classified as having a negative, neutral or positive sentiment or like in [41] where tweets are classified into categories such as News, Events, Opinions, Deals, and Private Messages. Not as much research has focused on classification of text as being radical/terrorist related or not.

One approach is described in [46] where radical tweets are classified in the categories Media, War terrorism, Extremism, Operations, Jihad, Country and Al-Qaeda using security dictionaries of enriched themes where each theme was categorized by semantically related words. In [46] they built dictionaries by looking at tweets containing hashtags like *Al-Qaeda, Jihad, Terrorism* and *Extremism* and by collecting relevant words for their purpose. A document was vectorized not according to the frequency of words but on the basis of presence of security related keywords. For example if the categories in which the messages should be classified are Jihad, Terrorism and Country, then for each category that the message contains related words the value will be 1 otherwise 0. The presence of one or more words relevant to predefined categories (War-Terrorism, Extremism, Jihad etc.) was used to deduce final category. The high results obtained, over 90% accuracy, led to the idea that keywords might be a good approach in classifying tweets.

An approach using ISIS related tweets to predict future support or opposition for ISIS was done in [36] where the authors used Twitter data to study the antecedents of ISIS support of users. As features vectors, they used bag of words features including individual terms, hashtags and user mentions. Bag of words is the representation of text as a multiset of its words. At a personal, historic level, they managed to predict future support or opposition of an user for ISIS with 87% accuracy. For this they trained a SVM classifier with a linear kernel with default parameters. One of the problems encountered in their work was to separate pro-ISIS tweets to con-ISIS tweets. They noticed that in anti-ISIS tweets when referring to Islamic State users write ISIS (77.3% of the tweets), while in pro-ISIS tweets they write Islamic State (93.1%). The good result of the classifier indicates that SVM might be a good approach in classifying tweets. In this work we are separating pro-ISIS tweets and tweets against ISIS. In this work we use a list of users that are divided into clusters with known ISIS supporters and therefore we did not have to use methods such as the one described in [36] to separate pro ISIS tweets from tweets against ISIS.

In [25] a study using sentiment analysis of tweets was conducted. They classify a tweet as being negative, neutral or positive. Some of the features are based on the polarity of words. This is determined by using several dictionaries like Dictionary of Affect in Language (DAL) or WordNet which assigns each word a pleasantness score between 1 (negative) and 3 (positive). Other features include counting features (counting the number of positive or negative words) and presence of exclamation marks and capitalized text. The polarity of words feature, counting and presence of exclamation marks and capitalized text form what they call senti-features. In their experiments using a SVM classifier and unigram features they get 71.36% accuracy. On the other hand when unigram features are combined with senti-features the result increases to 75.39% showing the contribution of the senti-features for tweets sentiment classification. Go et al. [27] have done another study in tweet sentiment classification. They used machine learning algorithms like Naive Bayes, maximum entropy, and SVM for classification. In their approach the emoticons (facial expressions pictorially represented using punctuation and letters usually used to represent the user's mood) are stripped out from the training data because there is a negative impact on the accuracies of the maximum entropy and SVM classifiers. This approach allows the classifiers to learn from other relevant feature they use like unigrams, bigrams or part of speech. Bigrams are used to classify tweets that contain negated phrases like "not good", or "not bad". In their experiments negation as an implicit feature with unigrams does not improve accuracy so they use bigrams as well. Compared to unigrams features, accuracy improves for Naive Bayes from 81.3% to 82.7%. Since bigrams seem to help in increasing the accuracy of classifying tweets we use them in this work as well.

Twitter provides a list of most popular topics people tweet about known as trending topics in real time but it is often hard to understand what these trending topics are about. In [33] Twitter trending topics are classified into 18 different categories like sports, politics, technology etc. A bag-of-words approach for text classification is used. For each topic, a document is made from trend definition and varying number of tweets. The $tf-idf$ (term frequency inverse document frequency) weights are computed for each word. The $tf-idf$ measure allows to evaluate the importance of a word(term) to a document. This, $tf-idf$ is used to filter out common words. For each of the 18 labels, top most 500 or 1000 frequent words with their $tf-idf$ weights are used to build the dataset for machine learning. The best accuracy are obtained from using Naive Bayes Multinomial classifier (65.36%). It performs better that Naive Bayes (45.31%) and SVM (61.76%).

4 Building the classifier

4.1 Datasets

In this work we used three different datasets. We will call these datasets TW-PRO, TW-RAND and TW-CON. The datasets are described in Table 1.

Dataset	Description
TW-PRO	Tweets that are pro ISIS, based on hashtags and network of known jihadists.
TW-RAND	Randomly collected tweets discussing various topics.
TW-CON	Tweets from accounts that are against ISIS.

Table 1: The datasets used for the experiments.

TW-RAND consists of 2000 random tweets. The topics discussed were varying and were not related to ISIS. An example of such a tweet is following one:

"RT @KimKardashian: I can't wait for Call Of Duty Black Ops II to come out!!!! The graphics look crazy"

TW-CON consists of tweets from accounts that were talking about ISIS and some of them were even against it but none of them supporting it. Examples of such accounts are: stopisisforever, No2ISISofficial, STOPISIS2GETHER, anti_isis_iraq. The assumption that these accounts were not posting messages supporting ISIS was made based on the user name and manual verification. Accounts with user names such as the ones mentioned above are most probable promoting messages against ISIS. Examples of tweets posted by those accounts are:

Example 1:

"Iraqi forces fight ISIS to recapture Tikrit http://video.foxnews.com /v/4088263981001/iraqi-forces-fight-isis-to-recapture-tikrit"

Example 2:

"@RudawEnglish http://bit.ly/1sEYQsg sign the #petition

to let #UN #US #help the Yezidis of #Kurdistan prevent #ISIS #genocide #StopISIS"

Both the datasets TW-RAND and TW-CON form negative cases (tweets with non-radical content). Beside negative cases we also need positive cases (tweets that contains radical content) to be able to build a classifier that can recognize radical tweets.

TW-PRO is the dataset containing tweets that support ISIS. Finding a dataset that contains tweets that are radical is a difficult task. The most common approach is to use humans that manually classify tweets as radical or non-radical. In this work we used another approach to find a suitable dataset that we can use to train our algorithms on. We have collected a set of tweets containing hashtags that were related to jihadists, and in particular ISIS, from the English language spectrum of pro-ISIS clusters on Twitter. All of the hashtags we used that are listed below have a corresponding Arabic hashtag and are often used within Arabic and non-Arabic tweets to widen the availability of ISIS material in general. We have focused on the English hashtags. The hashtags we have used to collect data are the following:

- #IS
- #ISLAMICSTATE
- #ILoveISIS
- #AllEyesOnISIS
- #CalamaityWillBeFallUS
- #KhalifaRestored
- #Islamicstate

Information that is available about a tweet can be found in Table 3.

It was also the case that not all the tweets were written in English and most important not all of the tweets were messages supporting ISIS. Some of the messages were not related to ISIS at all and they had no violent/radical content. They were inside the corpus because they contained some similar hashtags like #IS for example. In these cases the #IS hashtag was not referring to the Islamic State but to the verb "is" (to be). Some tweets containing hashtags referring to ISIS and actually talking about ISIS was removed because they contained messages that were against ISIS. The selection of tweets that were written in English was done by using [39].

When it comes to sorting text according to its meaning the problem is complex and requires semantic analysis. Our first approach to select the tweets that were only about ISIS and that are supporting ISIS was to create a bag of words related to terrorism and war. This method proved to not be very efficient since it was not possible to separate pro ISIS tweets from tweets against ISIS based on only the topic. Both kind of tweets contain terrorism and war related words and therefore the bag of words approach could not be used to differentiate the meaning of the sentence.

To tackle this issue we used a list of user accounts describing clusters of known Jihadist sympathizers (retweeting the same users, followers etc.). The list consisted of 6729 user names. At the end only tweets posted or retweeted by these users have been selected and used as positive cases. More information about the dataset TW-PRO that we used (containing pro-ISIS tweets) can be found in Table 2. All duplicate tweets were removed from the datasets.

Total number of tweets	36515
Number of duplicate tweets	0
Number of retweets	27464
Number of original tweets	9051

Table 2: TW-PRO dataset.

id	id of the tweet
created_at	when the tweet was created
text	the text of the tweet
user id	the user id
description	the description that the users provide about himself/herself
time_zone	the time zone
lang	the language set by the user

Table 3: Information that is available about a tweet.

In order to access the tweets easily they were stored in a database. First a parser was build using [3] that extracted the useful information from the json files containing our tweets. The database created has two tables: Tweet and User as shown in Table 4 and Table 5.

id	the id of the tweet
created at	when the tweet was created
text	the text of the tweet
in reply status id	the id of the tweet that was replied to
in reply user id	the id of the user that was replied to
in reply screen name	the screen name of the user that was replied to
is retweet	if a tweet is retweeted or not

Table 4: Database table for a tweet.

user id	the id of the user
user name	the user name of the user
location	the location set by user
description	the description oh the user
time zone	the time zone set by the user
language	the language set by the user

Table 5: Database table for a user.

A total of 135608 tweets were collected and stored. More details about the dataset is shown in Table 6.

Total number of tweets	135608
Number of duplicate tweets	3108
Number of accounts set in English	71719
Number of retweets	79737
Number of original tweets	55871

Table 6: All datasets.

All the datasets where preprocessed in a similar way as described in the following section.

4.2 Preprocessing the data

The data, tweets in our case, contains a lot of "noise" in its raw form. The "noise" consists of information that are not useful for machine learning models. Moreover, such noise can alter the accuracy of results and therefore preprocessing the dataset is necessary. To eliminating ”noise” and prepare the corpus for building the model the following preprocessing steps are done:

- Remove RT (retweet tag) and annotation(@username). For example a tweet like: ”RT @AkhbarMujahid3: #BREAKING New release by AlHayat Media ”Dabiq #3 A Call to Hijrah” https://t.co/vsLu10ZSIx #IS #Syria #Iraq” will be transformed to: ”#BREAKING New release by AlHayat Media ”Dabiq #3 A Call to Hijrah” https://t.co/vsLu10ZSIx #IS #Syria #Iraq”

- All URLs (i.e., tokens beginning with http or www) are removed. A tweet like: ”Pentagon: #US military’s bombing raids & other operations in #Iraq cost 7.5m a day http://t.co/rYMw711CPD #IS #ISIS http://t.co/8nvQVjFKJq” will become: ”Pentagon: #US military’s bombing raids & other operations in #Iraq cost 7.5m a day #IS #ISIS ”

- Html character codes (i.e., &...;) are replaced with an ASCII equivalent. When downloading tweets the html character code is rendered instead of the ASCII code. A tweet like ”Pentagon: #US military’s bombing raids & other operations in #Iraq cost 7.5m a day #IS #ISIS ”, after replacing the html code with the ASCII code, will be ”Pentagon: #US military’s bombing raids & other operations in #Iraq cost 7.5m a day #IS #ISIS ”

- Lemmatize the text. Lemmatization is often used in computational linguistics problems. It is a process that determines the lemma of a word [35].
  In English a word can have different inflected forms. For instance the word ’walk’ can be used as ’walked’, ’walking’, ’walks’. The base form of those words is ’walk’. This base form of the words is called lemma. For example a text like ”He was with us yesterday and now he is tired” after lemmatizing will transformed to ”He be with we yesterday and now he be tired”.
  For lemmatization the toolkit [37] was used.

- Tokenization. Tokenization is often used in lexical analysis. It is the process that splits a text up into words, phrases, symbols or other elements. this elements are called tokens and they are usually used for

further processing [28].
Tokenization is important in text processing because it allows to process each item separately. An example of text tokenization can be the sentence "I can't wait, come or go!" after tokenization will be transformed to token1: "I", token2: "can",token3: "not",token4: "wait",token5: ",",token6: "come",token7: "or",token8: "go",token9: "!". The corpus was tokenized by using [37] and adding some additional transformations.

4.3 Feature vectors

When creating the feature vectors we used three different set of features:

- stylometric features (S)
- time based features (T)
- sentiment based features (SB)

Those approaches combined produced 829 features as described in the following sections. The number of the different set of features can be seen in Table 7.

stylometry based features	811
time based features	37
sentiment	5

Table 7: Number of features.

4.3.1 Stylometric features

Stylometry looks at the variations of literary style between different writers. Usually it includes statistics about the frequency of specific items or the length of words or sentences.

A common stylometric analysis method is called writer invariant or author invariant. It claims that all texts written by the same author are similar or invariant. In other words texts written by the same author will be more similar than those written by different ones. Even though we are not interested in the authors of tweets, stylometry is an important analysis method, since the topic all jihadist authors write about is similar and the purpose of the messages is the same, mainly to spread ISIS propaganda, it is reasonable to believe that the style of writing they have might be similar. A common approach of writer invariant method is the frequency of function words. Beside frequency of function words, stylometric analysis can include length of sentences, the number of sentences etc.

A function word is a word that has little lexical meaning or ambiguous meaning and is used to link other parts of speech in a sentence. Function words features are commonly used in text recognition problems. In this work we focus only on stylometric statistics applied for words. This due to the fact that a tweet cannot be longer than 140 characters. In addition, the frequency of hashtags are analyzed. Table 8 contains the features used for the stylometric analysis.

function words	frequency of various function words	293
frequent words	frequency of most frequent words	173
punctuation	frequency of characters . , , , ; , : , ' ,- , [,] , { , } , !,?,&	13
hashtags	frequency of most frequent hastags	100
letter bigrams	frequency of most frequent letter bigrams	133
word bigrams	frequency of most frequent word bigrams	99

Table 8: Stylometric features.

The list of the function words, frequent words and hashtags that are used in the stylometric analysis can be found in the Appendix.

The stylometric analysis is done as follows:

- First a vector of the same size as the numbers of function words (293) is created. For each position in vector we associate a function word with 0.

function word1	function word2	....	function word293
0	0	....	0

- When a tweet is parsed and a function word is found in the tweet, the value corresponding to that function word in the vector is increased by 1.
- When the whole tweet is read, the corresponding vector is normalized, meaning that each number associated with is divided with the total number of distinct function words contained in the tweet.

An example of building a function word vector can be considered on the following tweet:

"Iraqi forces fight ISIS to recapture all Tikrit"

In the tweet and the list of function words we can find two common function words: *to* and *all*. We increase the corresponding number for theses words by 1. At this moment we have a vector that looks like following:

function word1	all	................to	....	function word292	function word293
0	1	....0....0... 1	...0.....0.....	0	0

The last step consists of normalizing the vector. Each number is divided with the number of distinct function words that are in the tweet and the list with 293 function words. In this example we have two such words. After normalizing the vector will have following values:

function word1	all	................to	....	function word292	function word293
0	0.5	....0....0... 0.5	...0.....0.....	0	0

Notice that at the end the sum of all values contained in the vector will always be 1. Messages that contain radical content that are posted on Twitter by those who support terrorist organizations like ISIS tend to follow similar topics. Therefore, we use the most frequent words from the dataset containing tweets against ISIS. The words that we have selected all have a frequency greater than 180. A total of 173 words have been selected. The process of creating the feature vector for most common words is identical with the one of creating function words feature vector described previously.

When selecting the most frequent words we ignored stop words. Stop words are commonly used words like conjunctions or prepositions. A total of 32 stop words was used. The list of stop words was created by selecting short function words from the function words list.

Hashtags are used to emphasize the meaning and importance of some words and to permit users to easily find messages with specific themes or content. Hashtags usually contain the essence of the message. Similar to the process of collecting the most frequent words in the pro-ISIS dataset, we collected the most frequent hashtags. Only hashtags that appeared in the dataset more than 75 times were collected. At the end we had a list of 100 hashtags. As expected the list contained hashtags about ISIS or their activities. For example the two most frequent hashtags used were #IS and #AllEyesOnISIS. The entire list of hashtags can be found in the Appendix section. The process of creating a feature vector for hashtags is once again identical to the one of creating function words vector or most frequent words vector.

Punctuation is another stylometric feature. Here, 13 different punctuation marks are considered. Even though a tweet is not a long text and therefore there are not many punctuation marks, users sometimes repeatedly use some punctuation marks like question or exclamation mark in order to emphasize wonder or excitement. An example of this usage is:

> *"Who wants the truth about ISIS??? Well here it it from Sheikh AlAdnani in his recent speech #AllEyesOnISIS http://t.co/LFT790b5bo"*

Since bigrams have been successfully used before to classify tweets [27] we decided to use bigrams. In this work we have two approaches for using bigrams. First one is to use word bigrams which means that a bigram will be formed by two words. The other approach is to use letter bigrams which means we considered forming a bigram with two letters. We only selected word bigrams that had the frequency greater than 250. At the end we formed a list containing 99 word bigrams. The same process was followed to create the list of letter bigrams. Only letter bigram with a frequency greater than 3000 were selected. We formed a list containing 133 letter bigrams. The process of forming the word bigram feature vector and letter bigram feature vector is the same as forming function words, most frequent words and hashtags feature vectors previously described.

4.3.2 Time

In this work we used different features that describes time aspect of the tweets. These features are listed in Table 9.

hour	the hour when the tweet was posted	24
day of week	the day of the week when the tweet was posted	7
period of week	when the tweet was posted: weekday/weekend	2
period of day	the period of day when the tweet was posted	4

Table 9: Number of features in the time vector.

We divided a day into 4 periods of 6 hours each starting from 00:00. Since the data was collected between June 2014 and August 2014 we did not consider using months as features. To build a feature vector for time, the following steps were done:

1. A vector of size 37 is created and filled with 0:s.
2. All vector items corresponding to the time when the tweet was posted are set to 1.

Consider a tweet that has the date:

Fri Aug 29 19:20:40 +0000 2014

In this case the features will be:

- hour: 19:00
- day of week: Friday
- period of week: weekday
- period of day: 3

and the feature vector will look like:

hour1	....	hour19	Monday	....	Friday	weekend	....	period3	period4
0		1	0		1	0		1	0

4.3.3 Sentiment

Sentiment analysis refers to the attitude of the writer towards a specific topic or a text. It has been used extensively to classify sentiments in movie and book reviews and tweets [25, 43]. In this work we have investigated if the radical content spread by users on Twitter has any correlation with the sentiment that is associated to the message. For this we used a sentiment analysis tool described in [37]. The sentiment analysis tool is mainly developed for predicting the sentiment of movie reviews. The tool is useful in our setting as well because it works not only by looking at each word separately but it builds up a representation of whole sentences based on the sentence structure. The sentiment is computed based on how words compose the meaning of the sentence. The values the sentiment can take are: very negative, negative, neutral, positive, very positive. In the next picture the live demo of the tool was used to see the representation of the tweet "#ISIS kills an will continue killing the crusades #AllEyesOnISIS" and its associated sentiment.

As can be seen in Figure 9 the computed sentiment of a tweet is negative. We computed the sentiment of all tweets from the training data set and the average sentiment was negative. Most of the tweets we consider have negative sentiments. The sentiment feature vector has the length of 5, where each feature corresponds to a sentiment. The vector was built following the same steps as in the previous examples.

4.3.4 Feature selections

In order to only select features that contribute to classify the tweets, information gain has been performed. Features that had no contribution have been removed. The initial structure of feature vectors can be seen in Table 7. The structure of feature vectors after features selection can be seen in Table 10.

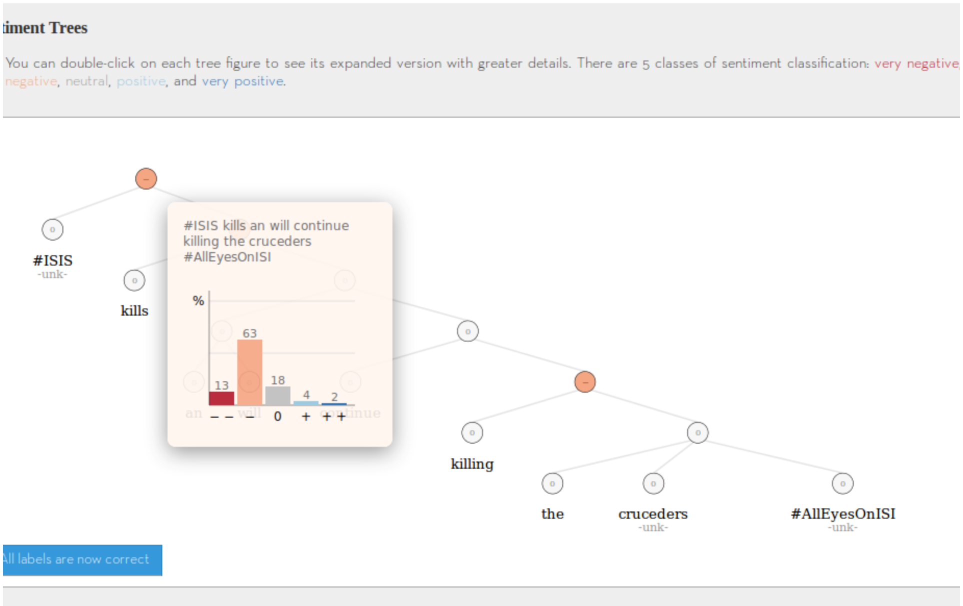

Figure 9: Sentiment tree representation of a tweet [13]

stylometry based techniques	579
time based techniques	36
sentiment	4

Table 10: The number of feature after the feature selection process.

5 Experiments

5.1 Experimental set-up

The experiments were conducted using a tool called Weka which is a suite of machine learning software written in Java. For our experiments we used three different classifiers: Support Vector Machine (SVM), Naive Bayes and AdaBoost. For Naive Bayes and AdaBoost the default configuration was chosen. For SVM a linear kernel was used.

5.2 Results

We performed experiments using different classification techniques, datasets, and features to evaluate the ability these techniques have to distinguish extremist from non-extremist tweets. We also examine which features appeared to be most important in facilitating this task. Between 4000 and 7500 tweets were used in total.

Results differed depending on what datasets we used. Using TW-PRO and TW-RAND led to better results than if TW-PRO and TW-CON were used. The results for TW-PRO and TW-RAND and the features (S + T + SB) are shown in Table 11. As can be noted AdaBoost performs very well with 100 % accuracy on the test set.

	Non Radical	Radical	Correctly Classified Instances
SVM	1974	24	99.1 %
	11	1990	
Naive Bayes	1997	1	99.9 %
	1	1990	
AdaBoost	1998	0	100 %
	0	1991	

Table 11: Results when using features (S + T + SB) on the datasets TW-RAND and TW-PRO.

The results for using the datasets TW-PRO and TW-CON are shown in Table 12, the accuracy when using the AdaBoost classifier is still high (99.5%). Since the datasets that are used for the experiments in Table 11 and Table 12 differs the results are expected. TW-RAND contains randomly selected tweets while TW-CON are tweets that are against ISIS. The tweets that are against ISIS contains similar hashtags and topics as the TW-PRO dataset and is therefore harder to separate than the randomly collected tweets.

	Non Radical	Radical	Correctly Classified Instances
SVM	1155	38	98.5 %
	24	2783	
Naive Bayes	1178	15	96.8 %
	114	2693	
AdaBoost	1182	11	99.5 %
	8	2799	

Table 12: Results when using all features (S + T + SB) on TW-CON and TW-PRO.

	Non Radical	Radical	Correctly Classified Instances
SVM	3099	92	97.9 %
	74	4813	
Naive Bayes	2877	314	89.0 %
	574	4313	
AdaBoost	1600	0	100 %
	0	2905	

Table 13: Results when using all features (S + T + SB) on the full dataset.

Table 13 shows the results for three different classifiers using all features on all the datasets TW-PRO, TW-RAND and TW-CON. As can be seen in the table AdaBoost performs slightly better than both Naive Bayes and SVM.

Given the importance of time features, we decided to examine the degree to which performance was compromised when these were excluded. Table 14 shows the result without time features.

	Non Radical	Radical	Correctly Classified Instances
SVM	3099	92	97.7 %
	88	4718	
Naive Bayes	2874	317	89.1 %
	550	4256	
AdaBoost	1576	0	100 %
	0	2423	

Table 14: Results when using all features except time (S + SB) on the full dataset.

Even without using time features, the results remain extremely impressive, with AdaBoost continuing to perform perfectly on the test data.

6 Conclusions

In this work we have presented an approach to classify tweets as containing radical content or not. There have been other attempts to classify radical content on Twitter. We used three different types of features: stylometry based features, time based features and sentiment based features. The results of the experiments proved that these features combined perform better than each individual one.

In order to have relevant results we used different datasets. For our testing dataset we collected random tweets and tweets having messages oriented against ISIS. For the training dataset, the pro ISIS tweets were collected from accounts that cluster with known ISIS supporters.

We run our experiments using three different classifiers: SVM, Naive Bayes and AdaBoost. The excellent results we obtained indicates that classification is a viable way forward to detect radical content on social media, and in particular on Twitter. We look forward to trying to replicate these results on more diverse and or complex data.

7 Future Work

In this work we covered many aspects regarding the attempt to classify radical content on Twitter. Still, there are many ways in which this work can be improved and extended.

It has been observed that radical tweets have a very low ageing factor (AF) [44]. It is a metric showing how fast a tweet was retweeted in a period of time. It is computed as follow:

$$AG = \sqrt[i]{\frac{k}{k+l}}$$

where i is the cut-off time in hours, k is the number of retweets originating at least i hours ago and l is the number of retweets originating less than i hours ago. A low AF value suggests by [45] that the topic is a short-term trending topic while a high value of the AF indicates that the topic is a sustainable topic since people have re-tweeted and discussed the tweet over a longer duration. The one hour ageing factor (1hAF) is the ratio of re-tweets in a sample set that originated more than one hour after the original creation time over the total number of re-tweets in the sample set.

The ageing factor plays an important role in our work due to the strategies that jihadist groups use to promote and promulgate messages. When a radical message is posted, those promoting such ideologies rush to re-tweet it. In the dataset that we have used in our experiments, the average 1hAF factor for a tweets is 0.06. This indicates that messages are re-tweeted quickly.

We believe that it will be interesting to use the ageing factor as part of our feature sets. One problem is that the perfect performance of our AdaBoost models makes it difficult to evaluate the relative importance of new features. More suitable tests will be possible once more complex data is found that we can do experiments on.

One way to improve the presented work is to make the classifier work not only on English tweets but also on tweets written in other languages. Since radical messages are not posted only in English, improving the classifier by making it work with tweets posted in other languages will be a significant contribution.

Another approach to improve this work is to focus not only on each tweet in particular to classify it as having radical content or not but also trying to label a Twitter account as being radical or not. The achieved result in this work might be the ground base for classifying users instead of tweets. The good results obtained in this work makes us optimistic about using the similar techniques to classify not only tweets but any other form of text.

8 APPENDIX

8.1 List of function words

1. a
2. able
3. aboard
4. about
5. above
6. absent
7. according
8. accordingly
9. across
10. after
11. against
12. ahead
13. albeit
14. all
15. along
16. alongside
17. although
18. am
19. amid
20. amidst
21. among
22. amongst
23. amount
24. an
25. and
26. another
27. anti
28. any
29. anybody
30. anyone
31. anything
32. are
33. around
34. as
35. aside
36. astraddle
37. astride
38. at
39. away
40. bar
41. barring
42. be
43. because
44. been
45. before
46. behind
47. being
48. below
49. beneath
50. beside
51. besides
52. better
53. between
54. beyond
55. bit
56. both
57. but
58. by
59. can
60. certain
61. circa
62. close
63. concerning
64. consequently
65. considering
66. could
67. couple
68. dare
69. deal
70. despite
71. down
72. due
73. during
74. each
75. eight
76. eigth
77. either
78. enough
79. every
80. everybody
81. everyone
82. everything
83. except
84. excepting
85. excluding
86. failing
87. few

88. fewer
89. fifth
90. first
91. five
92. following
93. for
94. four
95. fourth
96. from
97. front
98. given
99. good
100. great
101. had
102. half
103. have
104. he
105. heaps
106. hence
107. her
108. hers
109. herself
110. him
111. himself
112. his
113. however
114. i
115. if
116. in
117. including
118. inside
119. instead
120. into
121. is
122. it
123. its
124. itself
125. keeping
126. lack
127. less
128. like
129. little
130. loads
131. lots
132. majority
133. many
134. masses
135. may
136. me
137. might
138. mine
139. minority
140. minus
141. more
142. most
143. much
144. must
145. my
146. myself
147. near
148. need
149. neither
150. nevertheless
151. next
152. nine
153. ninth
154. no
155. nobody
156. none
157. nor
158. nothing
159. notwithstanding
160. number
161. numbers
162. of
163. off
164. on
165. once
166. one
167. onto
168. opposite
169. or
170. other
171. ought
172. our
173. ours
174. ourselves
175. out
176. outside
177. over
178. part
179. past
180. pending
181. per
182. pertaining
183. place
184. plenty
185. plethora
186. plus
187. quantities
188. quantity

189. quarter
190. regarding
191. remainder
192. respecting
193. rest
194. round
195. save
196. saving
197. second
198. seven
199. seventh
200. several
201. shall
202. she
203. should
204. similar
205. since
206. six
207. sixth
208. so
209. some
210. somebody
211. someone
212. something
213. sorry
214. spite
215. such
216. ten
217. tenth
218. than
219. thanks
220. that
221. the
222. their
223. theirs
224. them
225. themselves
226. then
227. thence
228. therefore
229. these
230. they
231. third
232. this
233. those
234. though
235. three
236. through
237. throughout
238. thru
239. thus
240. till
241. time
242. to
243. tons
244. top
245. toward
246. towards
247. two
248. under
249. underneath
250. unless
251. unlike
252. until
253. unto
254. up
255. upon
256. us
257. used
258. various
259. versus
260. via
261. view
262. wanting
263. was
264. we
265. were
266. what
267. whatever
268. when
269. where
270. whereas
271. wherever
272. whether
273. which
274. whichever
275. while
276. whilst
277. who
278. whoever
279. whole
280. whom
281. whenever
282. whose
283. will
284. with
285. within
286. without
287. would
288. yet
289. you
290. your
291. yours
292. yourself
293. yourselves

8.2 List of frequent words

1. state
2. islamic
3. not
4. soldier
5. do
6. kill
7. support
8. abu
9. allah
10. people
11. al
12. now
13. army
14. muslim
15. city
16. force
17. fight
18. control
19. take
20. against
21. iraqi
22. give
23. battle
24. say
25. destroy
26. village
27. isis
28. capture
29. distribute
30. syrian
31. province
32. regime
33. report
34. iraq
35. attack
36. fighter
37. assad
38. media
39. new
40. mujahideen
41. division
42. base
43. islam
44. today
45. come
46. join
47. poor
48. use
49. get
50. remove
51. military
52. show
53. release
54. war
55. brother
56. clash
57. under
58. video
59. make
60. area
61. group
62. see
63. flag
64. go
65. picture
66. free
67. storm
68. gas
69. time
70. caliphate
71. pay
72. mosul
73. allegiance
74. how
75. tax
76. field
77. world
78. just
79. leader
80. between
81. send
82. leave
83. training
84. street
85. only
86. spoil
87. aid
88. christian
89. pledge
90. call
91. assault
92. want
93. collect
94. operation
95. hold
96. day
97. need
98. militia
99. troops

100. part
101. security
102. israel
103. help
104. deir
105. here
106. bakr
107. parade
108. malikus
109. office
110. border
111. lie
112. resident
113. iranian
114. syrium
115. woman
116. via
117. distribution
118. another
119. hand
120. also
121. liberate
122. member
123. road
124. try
125. seize
126. account
127. back
128. year
129. soon
130. akbar
131. claim
132. weapon
133. start
134. last
135. carry
136. gain
137. defeat
138. follow
139. because
140. flee
141. u
142. vehicle
143. supporter
144. accept
145. photo
146. still
147. official
148. service
149. khilafah
150. cob
151. martyr
152. jihad
153. full
154. brigade
155. lion
156. anbar
157. gaza
158. burn
159. prisoner
160. live
161. caliph
162. ask
163. front
164. victory
165. dead
166. country
167. bomb
168. barracks
169. khilafa
170. death
171. name
172. land
173. issue

8.3 List of Hashtags

1. #IS
2. #AllEyesOnISIS
3. #Iraq
4. #Islamic_State
5. #Syria
6. #IslamicState
7. #Islam
8. #ISIS
9. #KhilafaRestored
10. #Muslims
11. #Mosul
12. #Khilafah
13. #Caliphate
14. #Jihad
15. #Raqqa

16. #Zakat
17. #Khilafa
18. #Gaza
19. #Baghdad
20. #Israel
21. #Homs
22. #Army
23. #BREAKING
24. #CalamityWillBefallUS
25. #Quran
26. #is
27. #Kirkuk
28. #Damascus
29. #army
30. #GazaUnderAttack
31. #Breaking
32. #Aleppo
33. #Ramadan
34. #Palestine
35. #iraq
36. #SAA
37. #Tikrit
38. #US
39. #Anbar
40. #Saudi
41. #Shia
42. #syria
43. #Iran
44. #URGENT
45. #islamicstate
46. #isis
47. #USA
48. #Diyala
49. #Haditha
50. #Muslim
51. #SaudiArabia
52. #Live_The_Cause
53. #PKK
54. #Lebanon
55. #Christians
56. #Indonesia
57. #Bayah
58. #Hasaka
59. #Sunnis
60. #Israeli
61. #AlHayat_Media
62. #Raqqah
63. #JN
64. #Hezbollah
65. #Syrian
66. #Jordan
67. #Islamicstate
68. #Iraqwar
69. #FSA
70. #Jews
71. #Sham
72. #Nigeria
73. #ISIL
74. #Pakistan
75. #Nineveh
76. #Kurds
77. #PRT
78. #Iranian
79. #khilafarestored
80. #New
81. #AQ
82. #Syri
83. #YPG
84. #Iraqi
85. #Khalifah
86. #AlHayat
87. #PT
88. #WorldCup2014
89. #UN
90. #Live_the_cause
91. #khilafah
92. #Salahadeen
93. #Racism
94. #Kashmir
95. #Assad
96. #orphans
97. #Maliki
98. #REPORT
99. #Hasakah
100. #saudi

References

[1] Confusion matrix. http://aimotion.blogspot.se/2010/08/tools-for-machine-learning-performance.html. [Online; accessed 15-April-2015].

[2] Currently listed entities. http://www.publicsafety.gc.ca/cnt/ntnl-scrt/cntr-trrrsm/lstd-ntts/crrnt-lstd-ntts-eng.aspx. [Online; accessed 20-June-2015].

[3] Encode or decode json text library for java. http://code.google.com/p/language-detection/.

[4] First classification of a weak learner. https://alliance.seas.upenn.edu/~cis520/wiki/index.php?n=lectures.boosting. [Online; accessed 19-April-2015].

[5] Foreign Terrorist Organizations. http://www.state.gov/j/ct/rls/other/des/123085.htm. [Online; accessed 20-June-2015].

[6] How Does ISIS Recruit, Exactly? Its Techniques Are Ruthless, Terrifying, And Efficient. http://www.bustle.com/articles/40535-how-does-isis-recruit-exactly-its/techniques-are-ruthless-terrifying-and-efficient. [Online; accessed 11-June-2015].

[7] ISIS hashtag campaign. http://www.telegraph.co.uk/news/worldnews/middleeast/iraq/10923046/How-Isis-used-Twitter-and-the-World-Cup-to-spread-its-terror.html. [Online; accessed 11-June-2015].

[8] ISIS Is Winning the Online Jihad Against the West. http://www.thedailybeast.com/articles/2014/10/01/isis-is-winning-the-online-jihad-against-the-west.html. [Online; accessed 11-June-2015].

[9] Machine learning flow. http://commons.wikimedia.org/wiki/File:Machine_Learning_Technique..JPG. [Online; accessed 10-April-2015].

[10] Mujahideen. http://terrorism.about.com/od/m/g/Mujahideen.htm. [Online; accessed 10-June-2015].

[11] Overfitting. http://www.analyticsvidhya.com/blog/2015/02/avoid-over-fitting-regularization/. [Online; accessed 25-April-2015].

[12] PROSCRIBED TERRORIST ORGANISATIONS. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/417888/Proscription-20150327.pdf. [Online; accessed 20-June-2015].

[13] Sentiment analysis. http://nlp.stanford.edu:8080/sentiment/rntnDemo.html. [Online; accessed 30-April-2015].

[14] The kernel trick. http://www.nelsonspencer.com/blog/2015/2/15/machine-learning-supervised-learning-pt-2. [Online; accessed 16-April-2015].

[15] The margin and support vectors. https://www.dtreg.com/solution/view/20. [Online; accessed 15-April-2015].

[16] The Non-separable case. http://docs.opencv.org/doc/tutorials/ml/non_linear_svms/non_linear_svms.html. [Online; accessed 15-April-2015].

[17] The separable case. http://docs.opencv.org/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html. [Online; accessed 15-April-2015].

[18] FBI. http://www.fbi.gov, 2014. [Online; accessed 10-March-2015].

[19] Internet Users. http://www.internetlivestats.com/internet-users/, 2014. [Online; accessed 10-March-2015].

[20] Isis propaganda: Study finds up to 90,000 Twitter accounts supporting extremist group. http://www.independent.co.uk/life-style/gadgets-and-tech/isis-propaganda-study-finds-up-to-90000-twitter/-accounts-supporting-extremist-group-10090309.html, 2014. [Online; accessed 10-March-2015].

[21] Security Council Adopts Resolution 2170. http://www.un.org/press/en/2014/sc11520.doc.htm, 2014. [Online; accessed 20-June-2015].

[22] Use of Internet for Terrorist Purposes. http://www.unodc.org/documents/frontpage/Use_of_Internet_for_Terrorist_Purposes.pdf, 2014. [Online; accessed 10-March-2015].

[23] Nico Prucha Ali Fisher. The call-up: The roots of a resilient and persistent jihadist presence on twitter ctx. 4(3). August 2004.

[24] Ethem Alpaydin. *Introduction to Machine Learning*. MIT Press, 2014.

[25] Ilia Vovsha Owen Rambow Rebecca Passonneau Apoorv Agarwal, Boyi Xie. Sentiment analysis of twitter data.

[26] Michael Collins, Robert E Schapire, and Yoram Singer. Logistic regression, adaboost and bregman distances. *Machine Learning*, 48(1-3):253–285, 2002.

[27] Alec Go, Richa Bhayani, and Lei Huang. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, pages 1–12, 2009.

[28] Peter Jackson and Isabelle Moulinier. *Natural language processing for online applications: Text retrieval, extraction and categorization*, volume 5. John Benjamins Publishing, 2007.

[29] Ali Fisher Jamie Bartlett. How to beat the media mujahideen. http://quarterly.demos.co.uk/article/issue-5/how-to-beat-the-media-mujahideen/, 2015.

[30] Finin T Tseng B Java A, Song X. Why we twitter: understanding microblogging usage and communities. *Proc of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, pages 56–65, 2007.

[31] Thorsten Joachims. *Text categorization with support vector machines: Learning with many relevant features*. Springer, 1998.

[32] John T Kent. Information gain and a general measure of correlation. *Biometrika*, 70(1):163–173, 1983.

[33] Kathy Lee, Diana Palsetia, Ramanathan Narayanan, Md Mostofa Ali Patwary, Ankit Agrawal, and Alok Choudhary. Twitter trending topic classification. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, pages 251–258. IEEE, 2011.

[34] David D Lewis. Naive (bayes) at forty: The independence assumption in information retrieval. In *Machine learning: ECML-98*, pages 4–15. Springer, 1998.

[35] Haibin Liu, Tom Christiansen, William A Baumgartner Jr, and Karin Verspoor. Biolemmatizer: a lemmatization tool for morphological processing of biomedical text. *J. Biomedical Semantics*, 3(3):17, 2012.

[36] Weber Ingmar Magdy Walid. #failedrevolutions: Using twitter to study the antecedents of isis support. *arXiv preprint arXiv:1503.02401*, 2005.

[37] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, Baltimore, Maryland, June 2014. Association for Computational Linguistics. http://www.aclweb.org/anthology/P/P14/P14-5010.

[38] Dorian Pyle. *Data preparation for data mining*, volume 1. Morgan Kaufmann, 1999.

[39] Nakatani Shuyo. Language detection library for java. http://code.google.com/p/language-detection/, 2010.

[40] Devin R Springer. *Islamic radicalism and global jihad*. Georgetown University Press, 2009.

[41] Bharath Sriram, Dave Fuhry, Engin Demir, Hakan Ferhatosmanoglu, and Murat Demirbas. Short text classification in twitter to improve information filtering. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, pages 841–842, New York, NY, USA, 2010. ACM. http://doi.acm.org/10.1145/1835449.1835643.

[42] Stephen V Stehman. Selecting and interpreting measures of thematic classification accuracy. *Remote sensing of Environment*, 62(1):77–89, 1997.

[43] Tun Thura Thet, Jin-Cheon Na, and Christopher SG Khoo. Aspect-based sentiment analysis of movie reviews on discussion boards. *Journal of Information Science*, page 0165551510388123, 2010.

[44] Victoria Uren and Aba-Sah Dadzie. Ageing factor: a potential altmetric for observing events and attention spans in microblogs. In *1st International Workshop on Knowledge Extraction and Consolidati on from Social Medi(KECSM)*, 2012.

[45] Victoria Uren and Aba-Sah Dadzie. Nerding out on twitter: Fun, patriotism and #curiosity. In *Proceedings of the 22Nd International Conference on World Wide Web Companion*, WWW '13 Companion, pages

605–612. International World Wide Web Conferences Steering Committee, 2013.

[46] Pooja Wadhwa and M . P . S . Bhatia. *Case Studies in Secure Computing Achievements and Trends*, chapter Classification of Radical Messages on Twitter Using Security Associations, page 273. 2014.

[47] Ian H Witten and Eibe Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.

[48] Dan Zarrella. *The social media marketing book*. " O'Reilly Media, Inc.", 2009.

[49] Tong Zhang. An introduction to support vector machines and other kernel-based learning methods. *AI Magazine*, 22(2):103, 2001.

[50] Xiaojin Zhu and Andrew B. Goldberg. *Introduction to Semi-Supervised Learning*. 2009.