# Machine Learning and Social Media in Crisis Management: Agility vs Ethics

**Dr. Vitaveska Lanfranchi**
OAK Group
Department of Computer Science
University of Sheffield
v.lanfranchi@sheffield.ac.uk

#### ABSTRACT

One of the most used sources of information for fast and flexible crisis information is social media or crowdsourced data, as the information is rapidly disseminated, can reach a large amount of target audience and covers a wide variety of topics. However, the agility that these new methodologies enable comes at a price: ethics and privacy. This paper presents an analysis of the ethical risks and implications of using automated system that learn from social media data to provide intelligence in crisis management. The paper presents a short overview on the use of social media data in crisis management to then highlight ethical implication of machine learning and social media data using an example scenario. In conclusion general mitigation strategies and specific implementation guidelines for the scenario under analysis are presented.

#### Keywords

Machine Learning, Social Media, Intelligent systems, Ethics, Privacy, Mitigation Strategies.

#### INTRODUCTION

In recent years the amount of emergencies and disasters of various types and duration has been constantly increasing, (see the infographic published by The United Nations Office for Disaster Risk Reduction in Figure 1) and this trends is only going to progress. Designing new services for resilience is now a fundamental requirement: new agile methodologies for emergency management need to be implemented, that are faster and more flexible in identifying challenges, and sourcing and implementing new solutions. Answering to this need, several new methodologies such as crowdsourcing, citizen sensing and sensor technologies have been trialled in recent years (Culotta, 2010; Abel et al, 2012; Meier, 2015; Mazumdar, 2016), to ensure information about a disaster is rapidly acquired and shared, thus allowing faster and flexible reaction to the variety of emergencies.

The most used sources of information for fast and agile crisis information are probably social media or crowdsourced data, as information is rapidly disseminated, can reach a large amount of target audience and covers a wide variety of topics. Information on social media, however, is often i) duplicated, incomplete, imprecise, or incorrect (in some cases, deliberately so); ii) written in informal style (i.e., short, unedited and conversational), thus much less grammatically bounded and containing extensive use of shorthand, symbols (e.g., emoticons), misspellings, etc.; iii) generally concerning the short-term zeitgeist; and iv) covering every conceivable domain. Given the scale of social media and the characteristics of the information, automated methods to monitor and capture the information in social streams are often required. To this extent many new tools have been created to exploit social media and crowdsourcing during emergency, many of them adopting automatic or semi-automatic Machine Learning (ML) algorithms to process information. However, the combination of social data and machine learning algorithms to understand and filter the data has also high ethical implications, as it can introduce several types of bias (i.e. data collection bias, training set bias) during the creation of the models, thus potentially causing misapplication of models and flawed interpretation of results.

Whilst previous papers have analysed the ethics implications of using social media data, in this paper we focus on the implications of using machine learning-based tools (or other automated systems) that learn from social media or crowdsourced data to provide intelligence. We will first of all present a brief state of the art of social media analysis systems for emergencies and of the ethics risks of using social media data for emergencies. We will then explain how the introduction of ML algorithms to analyse the data can introduce new risks and we will discuss the mitigation strategies that can be put in place to ensure a respectful, ethical and effective usage of this channel.

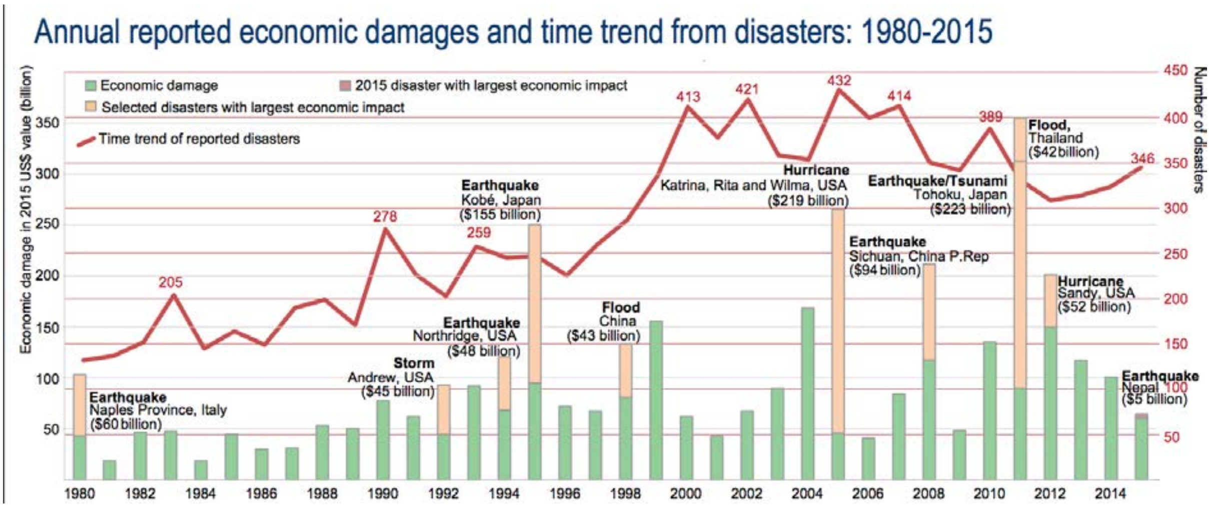


**Figure 1 - The United Nations Office for Disaster Risk Reduction disaster trend from 1980 to 2015 (source: http://www.unisdr.org/files/47804_2015disastertrendsinfographic.pdf)**

## SOCIAL DATA FOR EMERGENCIES – STATE OF THE ART

The use of social media data and crowdsourcing for emergencies has emerged a few years ago, when social media started becoming a major channel to disseminate and share information. Given the vast amount of information on social media and on crowdsourcing applications, finding relevant information in an effective manner has proved problematic: too much information can cause information overload and inability to separate relevant information from noise. To this extent, researchers have worked to create algorithms that analyse the information using NLP (Natural Language Processing) or ML technologies and provide either visual analytics solution to prioritise information, or algorithms for threat detection etc.
Existing systems for the analysis of social media or crowdsourced data vary accordingly to the type of analysis that is performed (e.g. offline or online) and to the intended outcome (e.g. visualization of data or recommendations) but they share the same conceptual idea of using Machine Learning and Information Extraction to extract information from social media streams and use that information to learn patterns and trends and to improve the algorithms.

Various studies have been performed on the possibility of social media data to be a predictor for public health, with research into Twitter Analysis to recognize early warnings for Swine Flu pandemic (Quincey and Kostkova, 2010), Dengue outbreaks (Gomide et al., 2010) and influenza (Culotta, 2010; Lampos, 2010). Another domain where social media data has been extensively researched is natural disasters, for example earthquakes: Sakaki et al in 2010 analysed Twitter data to propose a new algorithm for earthquake detection [Sakaki, 2010]; Caragea et al in 2011 proposed a system that uses Machine Learning (ML) for categorizing Tweets for the Haiti earthquake (Caragea et al, 2011).
Real-time analysis of social media data is often provided by tools that make use of visual analytics to visualize information and trends for users. Twitcident (Abel et al, 2012) uses ML to analyse Twitter Streams in real-time, filtering the Tweets accordingly to emergency categories and providing a set of visualization widgets to see the data. TRIDS is a system for monitoring social media that enables situation awareness in localised events, using Information Extraction techniques to analyse the data in real-time and visualize them using multiple faceted widgets (Ireson et al, 2015). SensePlace2 (MacEachren, 2011) is a geovisual system that using NLP techniques to extract location and time from Tweets and map them accordingly. SocialSensor is a EU project that focuses on real-time analysis of multimedia data streams, developing tools to support journalists and citizens (Papadopoulos et al, 2014), with a focus on emergencies (Papadopoulos et al, 2013).

Whilst social media streams provide large volume of data, the data available may be inconsistent, incoherent and not trustworthy. For this reason research effort in crisis management has focused on crowdsourcing, that is the process of enabling "capable crowds to participate in various tasks, from simply 'validating' a piece of information or photograph to complicated editing and management" (Gao et al, 2011). Crowdsourcing platforms have been used in many emergencies to collect information about damages (Yang et al., 2014) or have been used on a daily base to prevent emergencies by collecting sensor information about natural resources and alerting contributors when values are out of the normal range (Mazumdar et al, 2016). A different approach that combines crowdsourcing and social media data analysis has been proposed by Meier (2015), using volunteers' efforts to review automatically categorised social media data.

Whilst these approaches have been proven effective in real-life scenario, there is no doubt that aggregating, fusing, analysing and visualising information can exacerbates the privacy and ethics issues that are already present when dealing with "simple" social media or crowdsourced data (Watson et al, 2013).

### ETHICAL RISKS OF SOCIAL MEDIA ANALYSIS – STATE OF THE ART

Unintended consequences of Social media usage in emergencies situations have been analysed in various papers, with examples that show how social media has created an 'unintended "do-it-yourself" society' (Rizza et al, 2012) where social media are used as means of surveillance, without concern for the privacy rights of individuals that may be involved in an emergency. The famous case of identification of a wrong person as suspect for the Boston bombing and the consequences this has been investigated by Tapia et al (2014).
Other research in the area has identified the issues given by the fact that social media users may share information about other people without preserving their privacy: for example Watsons et al (2014) discuss issues raised by people sharing images of victims of the 2005 London Bombing. The rapidity of sharing information on social media means that, if a photo is removed as too sensitive or infringes someone's privacy, it may already be replicated on other channels.

### MACHINE LEARNING AND SOCIAL MEDIA FOR EMEWRGENCY MANAGEMENT

Machine Learning is a disciplines that studies, designs and develops algorithms that learn from experience. An example of ML for Emergency Management is the use of facial recognition embedded in CCTV cameras: facial recognition is often used as allows to quickly process millions of faces and recognise similar faces across multiple images but has also been proven as potentially racially and sex-biased (Klare et al, 2012; O'Toole et al, 2011). To better understand the potential ethics issues of applying ML techniques to Social Media it is important to understand how ML Algorithms work. The following image, taken from a Google Research presentation on ML, shows a basic diagram (Figure 2).

A Machine Learning algorithm works by having a set of example to learn from that are inputted into a model. The model is created by selecting a set of variables or factors that are used to make a judgment (prediction or identification) when a new data is presented to the model. A Machine Learning algorithm has also a learner component, that looks at differences between the judgment of the model and the actual outcome or truth to adjust the parameters and in turn the model.

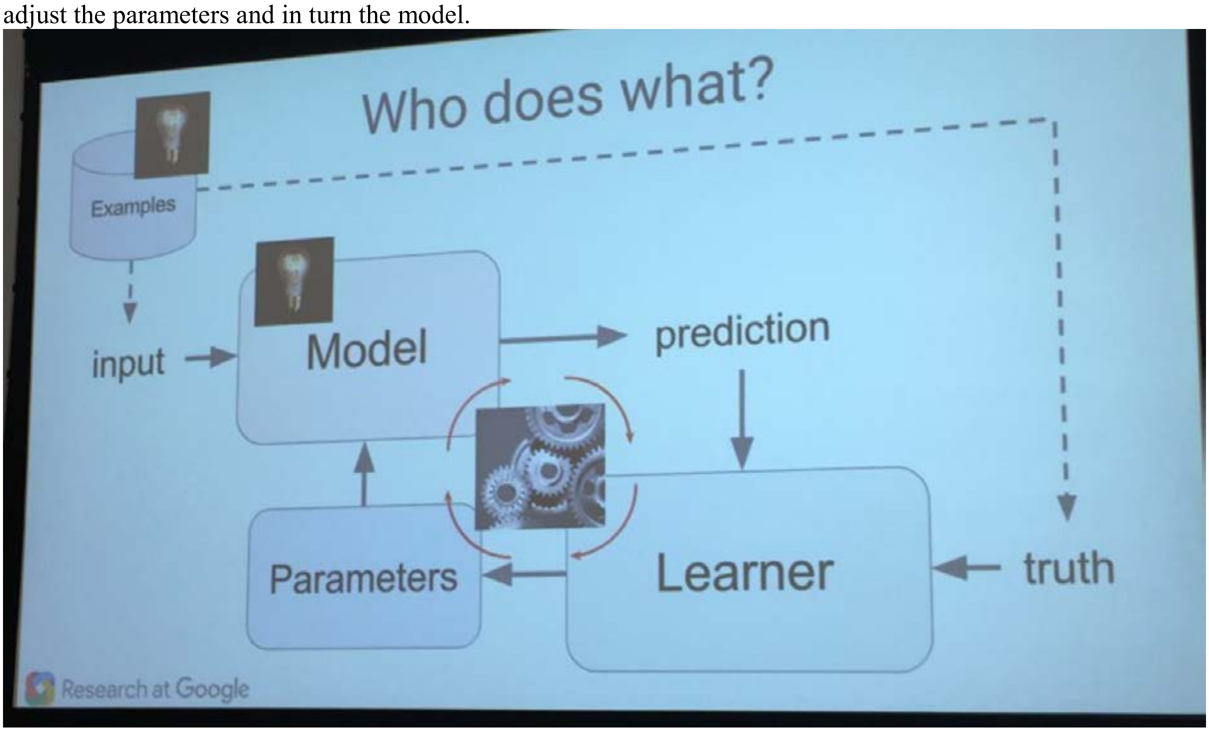

**Figure 2 – Google Research Diagram of Machine Learning (source: https://martechtoday.com/machine-learning-day-at-google-150275)**

Learning may occur in different manners, for example be supervised, unsupervised, semi-supervised etc. Given the nature of the algorithm process, there is a high risk that a machine learning system can reproduce patterns of discrimination and even exacerbate them (Barocas et al, 2016). Very often machine learning algorithms are "black-boxes" that output results, giving no visibility to the users to how the outcome was reached. Expert users may get visibility of an algorithm working process when analyzing the outcome, but again the visibility is depending on the type of algorithm used: for example, complex neural network algorithms have much less transparency than algorithms based on decision trees or Bayesian networks (Bostrom et al., 2014).

Bias can be introduced in a machine learning systems at different stages on the design and development process. A study by Tiell et al (2016) identified three stages of ethics bias risks when designing and developing a machine learning algorithm:

- ➔ System Design
    - Human Cognitive bias: this is a bias in cognitive processes (identifying, reasoning, abstracting evaluating, etc.), due to the personal preferences and beliefs of an individual (in this case, the system designer).
    - Algorithm selection: this bias occurs as the system designer will make a choice between different algorithms to analyse the data, therefore inheriting bias and risks associated with the algorithm chosen.
    - Data collection bias: this occurs when the data chosen is biased, for example because of non-random selection.
    - Missing or misquantified data: this occurs when data is missing, due to the way data is captured and or/manipulated/processed (e.g. labelled).
- ➔ Modeling and training:
    - Reinforcement bias: this occurs when selective thinking is applied, thus selecting and choosing only the data that confirms the researcher beliefs and ignoring the contradictory data.
    - Societal bias: this refers to the attitudes or stereotypes ingrained in the researcher's culture and education that can affect understanding and selection.
    - Safety boundaries: i.e. the mental maps of risk and safety that can influence choices.
    - Under-representation of minority classes: this is a bias due to the fact that certain type of data or systems may not represent all the cultures/minorities
    - Validation of data labels: this occurs when data labels are validated either by a human or by a system.
- ➔ Presentation and Implementation
    - Flawed interpretation of results: this occurs when there is voluntary (e.g. by the researcher) or involuntary (e.g. given by the fact that the system is flawed or by procedural / habitual behaviours in interpreting the results) manipulation in the **analysis** or reporting of findings
    - Misapplication of models: this occurs when the models are applied to the wrong data/domain

In the following section we will present a case study scenario where Social Media Data is collected for an emergency, analysed using a machine learning algorithm and the results are presented to the user using a visual interface.

### A Scenario

An emergency management team decides to use social media to understand the events related to the terrorist attack in Istanbul. They use a system that collects social media data legally and in line with the terms and conditions of Twitter. The Twitter data they gather will be fully identifiable.

The system gathers data using hashtags such as: #Turkey #Istanbul #ISIS #IS. The system will:

- perform sentiment analysis
- perform social network analysis
- create a network visualization

- starts following users that post relevant content

The aim is to understand how sentiment about the events emerges over time amongst different networks of Twitter user and identify specific users that show extreme emotions and can be, for the content and the sentiment shared, and their social network position, be identified as "threats".

In order for the system to achieve its aims, it needs to run some Machine Learning algorithms. For example, let's take Sentiment Analysis. In order to perform Sentiment Analysis, the system will have to be pre-trained on an existing corpus of Twitter data, ideally related to Emergency Management. The training corpus could be old (therefore obsolete) and chosen with a bias (e.g. only Tweets in English related to terrorism and a few other natural disasters): in the training corpus researchers will have annotated each Tweet with a sentiment, thus potentially introducing new bias (e.g. sentiment could be very subjective and culturally different). The ML learning algorithm will use that corpus to create a model, based on rules (e.g. "when the Tweet contains the words 'I am upset' the sentiment is negative") and will use this rules to automatically classify the incoming Tweets into positive, negative or neutral.

In the following table ( Table 1) we identify the possible ethics risk associated with the above scenario using the classification provided by Tiell et al. and other risks identified (for the purpose of this analysis we will limit the ethics risks to those associated with the interaction between machine learning and social media, not taking into account more generic risks associated with the use of social media data).

| **ETHICAL RISK** | **IMPLICATIONS** |
|---|---|
| Data collection bias. | Data is not representative of population that does not use social media or uses a different social media from the one concerned. |
| Missing or Erroneous Data | The system collects social media data using hashtags. Not everyone uses hashtags but may still use related terms therefore potentially missing relevant data.<br>Intelligent solutions that use content analysis and synonyms for gathering content allow for more flexibility and wider coverage but at the same time can bias the data towards the system "dictionary". |
| Misquantified or misrepresented data | The system will make judgements about tweets (i.e. sentiment) that may result in data being misrepresented as the true sentiment may not come across (e.g. sarcasm). |
| Reinforcement bias | When the system learns that a user is relevant (given the topic and the sentiment shown) it will start following that user to gather more content, thus potentially reinforcing the bias as more data is available from users highlighted as "relevant"<br>Rumours and false information, often spread intentionally by users trying to manipulate the system, may be reinforced. |
| Under-representation of minority classes and societal bias | Data is not representative of population that does not use social media and this will introduce an under-representation of minority classes. Moreover sentiment analysis systems are language dependent and if a sentiment analysis system is not able to cope with a specific language/dialect this will introduce a societal bias. |
| Validation of data labels | Given the high variety of topics and domains discussed on social media, the training set used as an input by the system may not be the most relevant for the topic under investigation. If the validation of the training test has been done automatically, this will introduce a biased linked to the fact that another |

| | |
|---|---|
| | algorithm has been used; if it is done manually it will be biased due to human factors. Mislabeling of training data can generate or exacerbate other biases.<br>Rumours and false information can be taken as truth by the system and used to train its algorithm. |
| Flawed interpretation of results | The system interpretation of data will be presented in a graphical form, thus potentially distracting the user from whether the data is misrepresented. |
| Data validity/availability bias | On Social Media, users or social media providers can remove information they posted, simply because they changed their mind or because they realise it is dangerous information or it is possibly infringing privacy regulations.<br>Given the fact that the system collects data from social media providers and stores it for analysis, data that has been later on deleted can be still visualized/available in the system.<br>Moreover deleted may be still part of the training set, therefore biasing the system judgement using data that should not be available. |

**Table 1. Ethical Risks in Machine Learning and Implications for Social Media Analysis**

## MITIGATION STRATEGIES

Mitigation strategies for the risks highlighted before can be introduced in a social media analysis system using a privacy by design framework, where all the privacy and ethics risks are considered and appropriate mitigation strategies for the case study are implemented.

The main mitigation strategies we have identified are:

- being *transparent*
- being *interactive*
- being *robust against manipulation*
- being *reactive*

The main mitigation strategy to ensure an ethical approach to Machine Learning for Social media data is to have a *transparent algorithm*, that provides users with means to monitor and understand the system internal working. Having a transparent algorithm will enable users to spot biases or mistakes and adopt corrective actions. For example, whilst data collected on social media may not be representative of all society, methods for quantify and control for selection bias can be adopted, using demographic inference techniques and looking at the information provided weighted by demographic (Culotta, 2014) to be aware, if not necessarily adjust for the bias. Corrective actions can also be taken to cope with missing data: various mitigation strategies exist ranging from using statistical methodologies to capture generic trends to develop remedial techniques (Robins et al., 2004; Sadikov et al, 2011).

Another important strategy is to have an *easy, flexible, interactive interface* that supports the user in investigating the data and provides full trace back of information a conclusion is based upon, with supporting evidence. Ideally an interactive interface will also provide means to interact with the algorithm, to report false positives or false negatives and improve the system performance.

To support ethics and privacy a machine learning system for social media should be *robust against manipulation*. For example, a sentiment analysis system for social media should be able to identify sarcasm, as sarcasm is a factor that can completely change the polarity of a sentence (Maynard et al, 2014). Being able to understand the sentiment and nature of a social media message very important: in 2010 a UK citizens was arrested under the Terror Act as authorities monitoring Social Media spotted a message he sent on Twitter regarding an airport (http://www.independent.co.uk/news/uk/home-news/twitter-joke-led-to-terror-act-arrest-and-airport-life-ban-1870913.html). This is a clear case where a police force monitored Social Media for terror relevant messages and flagged up as highly suspicious a sarcastic message. Another example of robustness against manipulation is the ability to cope with rumours and false information. In order to do so, multiple solutions could be employed, from a user verification and trustworthiness algorithm to multiple sources verification of a news. The sources used to verify a news could be official sources or could alternatively be other

Social Media users. In 2010 Mendoza et al published a work following the propagation of 'confirmed truths' and 'false rumors' on Twitter after an earthquake in Chile. Mendoza, Poblete, & Castillo, 2010 followed the propagation 'confirmed truths' and 'false rumors' on Twitter after an earthquake in Chile. They found that approximately 95.5% of tweets validated the 'confirm truths', and only 29.8% validated the 'false rumors'; while more than 60% denied or questioned them (Mendoza et al., 2010).

Finally, the system must be *reactive*, i.e. able to quickly act in response to a situation. For example, when a social media post is deleted, the system must ensure that the message is deleted from all its storage and from its training set, and rules learnt on the basis of that message should be "flagged" as potentially biased and needing confirmation. To avoid the risk of using "deleted" data in the learning and predicting process, the system should run on live data.

Whilst we have highlighted above generic mitigation strategies, it is useful to go back to our case study scenario to see how those mitigation strategies should be implemented in practical terms.

| **MITIGATION STRATEGY** | **POSSIBLE IMPLEMENTATIONS** |
|---|---|
| Data is not representative of population that does not use social media or uses a different social media from the one concerned. | The system should make use of the metadata available and of demographic information to allow the user to view information faceted by parameters, i.e. Based on location, demographic information etc. |
| The system collects social media data using hashtags. Not everyone uses hashtags but may still use related terms therefore potentially missing relevant data.<br>Intelligent solutions that use content analysis and synonyms for gathering content allow for more flexibility and wider coverage but at the same time can bias the data towards the system "dictionary". | The system should have powerful search operators to enable capturing wider content than hashtag, provide users with means to constantly monitor the used synonyms and keywords, provide means to easily remove keywords from the search. |
| The system will make judgements about tweets (i.e. Sentiment) that may result in data being misrepresented as the true sentiment may not come across (e.g. Sarcasm). | The sentiment analysis should be, if possible, customized for the domain, and offer a range of sentiments instead of a simple polarity. Advance techniques for sarcasm and irony detection should be adopted. |
| When the system learns that a user is relevant (given the topic and the sentiment shown) it will start following that user to gather more content, thus potentially reinforcing the bias as more data | The user interface should allow visibility of evidence (i.e. Tweets and networks relationship to illustrate why a user is considered relevant).<br>The user interface should provide users with |

| | |
|---|---|
| is available from users highlighted as "relevant" Rumours and false information, often spread intentionally by users trying to manipulate the system, may be reinforced. | means to constantly monitor the followed users and correct the monitoring if needed (e.g. Unfollow a user). The user interface can provide means to manually annotate message to reflect human analysis (e.g. Annotate a user as a divulger of false information). |
| Data is not representative of population that does not use social media and this will introduce an under-representation of minority classes. Moreover sentiment analysis systems are language dependent and if a sentiment analysis system is not able to cope with a specific language/dialect this will introduce a societal bias. | The user interface should clearly highlight if any data is missing from a representation, for example if sentiment analysis is available only for English messages but the system collects all languages, the user should be made aware of the bias. Simulations should be run on multiple datasets to determine whether the same results are produced for different populations or scenarios. |
| Given the high variety of topics and domains discussed on social media, the training set used as an input by the system may not be the most relevant for the topic under investigation. If the validation of the training test has been done automatically, this will introduce a biased linked to the fact that another algorithm has been used; if it is done manually it will be biased due to human factors. Rumours and false information can be taken as truth by the system and used to train its algorithm. | The internal workings of the algorithm should be explained to the user and the user should be able to provide custom training material for the algorithm. Possibility to feedback to the learning module should be provided to improve the system's effectiveness. |
| The system interpretation of data will be presented in a graphical form, thus potentially distracting the user from whether the data is misrepresented. | The user interface should choose visualisations that highlight the data dimensions but do not hide any missing data or misrepresent the available data. |
| On social media, users or social media providers can remove information they posted, simply because they changed their mind or because they realise it is dangerous information or it is possibly infringing privacy regulations. Given the fact that the system collects data from social media providers and stores it for analysis, data that has been later on deleted can be still visualized/available in the system. Moreover deleted may be still part of the training set, therefore biasing the system judgement using data that should not be available. | The system should, whenever possible, perform checks on the validity of the information it stores. For example, when receiving a notification from the Twitter API that a tweet is deleted, the system should delete the tweet from its internal storage. |

**Table 2 - Mitigation strategies for Machine Learning risks and Possible implementations**

## CONCLUSION

This paper started by reflecting on the use of social media during emergencies and how, given the large amount of data available and the need for fast, flexible, agile reaction, intelligent systems that categorise and analyse social media content are often used. However, the combination of social data and automatic or semi-automatic Machine Learning technologies to understand and filter the data can exacerbate existing privacy and ethics risk related to the use of social media in emergencies and also introduce new ones. To better understand how to develop a new generation of Machine Learning-based social media analysis tools to support emergency management, an analysis of the risks and mitigation strategies has been performed, using as an example a case study scenario of sentiment analysis on Social Media. This examination will be used as part of the iTrack project to guide the design and development of a solution for social media analysis in iTrack.

## ACKNOWLEDGMENTS

This work has been carried out as part of the iTrack project. The iTrack project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 700510".

## REFERENCES

Abel, F., Hauff, C., Houben, G., Stronkman, R. and Tao, K. (2012) Semantics + filtering + search = twitcident. exploring information in social web streams. In Proceedings of the 23rd ACM conference on Hypertext and social media (HT '12). ACM, New York, NY, USA, 285-294.

Barocas, S., and Selbst, A.D. (2016) Big data's disparate impact. California Law Review, 104:671–732, 2016.

Bostrom, N. and Yudkowsky, E. (2014) 'The ethics of artificial intelligence', in Frankish, K. and Ramsey, W.M. (eds.) *The Cambridge Handbook of Artificial Intelligence:*. Cambridge: Cambridge University Press, pp. 316–334.

Caragea, C., Giles, L., Jaiswal, A., Jansen, B.J., Kim, H., Mcneese, N., Mitra, P., Traylor, G., Tapia, A.H., Wu, D., & Yen, J. (2011). Classifying Text Messages for the Haiti Earthquake.

Culotta, A. (2010) Towards detecting influenza epidemics by analyzing twitter messages. In Proceedings of the First Workshop on Social Media Analytics, pages 115–122. ACM.

Culotta, A. (2014) Reducing Sampling Bias in Social Media Data for County Health Inference. JSM Proceedings.

Gao, H., Barbier, G., Goolsby, R. (2011). Harnessing the Crowdsourcing Power of Social Media for Disaster Relief. IEEE INTELL SYST APP, 10-14.

Gomide, J., Veloso, A., Meira, W., Almeida, V., Benevenuto, F., Ferraz, F., and Teixeira, M. (2011) Dengue surveillance based on a computational model of spatio-temporal locality of twitter. In Proceedings of the 3rd International Web Science Conference (WebSci '11). ACM, New York, NY, USA.

Hastie, T., Tibshirani, R. and Friedman, J. (2001). The Elements of Statistical Learning. New York, NY: Springer Science.

Ireson N., Lanfranchi, V., Mazumdar S., Ciravegna F. (2015) TRIDS: Real-time Incident Monitoring with Social Media (Inproceeding). In proceedings of Semantics and Analytics for Emergency Response (SAFE) workshop at ISCRAM, 2015.

Klare, B., Burge, M., Klontz, J., Vorder Bruegge, R., and Jain, A. (2012). Face Recognition Performance: Role of Demographic Information. *Trans. Info. For. Sec.* 7, 6 (December 2012), 1789-1801.

Kossinet, G (2006) Effects of missing data in social networks Social Networks, Vol. 28, No. 3. (July 2006), pp. 247-268.

Lampos, V. and Cristianini, N., (2010). Tracking the flu pandemic by monitoring the social web. In 2010 2nd International Workshop on Cognitive Information Processing (pp. 411-416). IEEE.

MacEachren, A.M., Jaiswal, A., Robinson, A. C., Pezanowski, S., Savelyev, A., Mitra, P., Zhang, X., Blanford, J. (2011) SensePlace2: GeoTwitter Analytics Support for Situational Awareness. IEEE Conference on Visual Analytics Science and Technology, Providence, RI, IEEE.

Mazumdar, S., Lanfranchi, V., Ireson, N., Wrigley, S., Bagnasco, C., Wehn, U., McDonagh, R., Ferri, M., McCarthy, S., Huwald, H., and Ciravegna, F,. (2016). Citizens observatories for effective Earth observations: the WeSenseIt Approach.Environmental Scientist journal, 2016

Maynard, D., Greenwood, M. (2014) Who cares about Sarcastic Tweets? Investigating the Impact of Sarcasm on Sentiment Analysis., In Proceedings of LREC.

Meier, P (2015). Digital humanitarians.

Mendoza, M., Poblete, B. and Castillo, C. (2010). Twitter under crisis: can we trust what we RT?. In *Proceedings of the First Workshop on Social Media Analytics* (SOMA '10). ACM, New York, NY, USA, 71-79.

O'Toole, A.J., Phillips, P.J., An, X., Dunlop, J. (2011) Demographic effects on estimates of automatic face recognition performance, Proceedings, Ninth International Conference on Automatic Face and Gesture Recognition, 2011.

Papadopoulos, S., Schinas, E., Mironidis, T., Iliakopoulou, K., Spyromitros-Xioufis, E., Tsampoulatidis, I. and Kompatsiaris, Y. (2013) Social Multimedia Crawling and Search

Papadopoulos, S., Kompatsiaris Y. (2014) Social Multimedia Crawling for Mining and Search. In Computer, vol.47, no.5, pp.84,87, May 2014.

Quincey, E. and Kostkova, P. (2010) Early warning and outbreak detection using social networking websites: The potential of twitter. Electronic Healthcare, pages 21–24.

Rizza, C., Guimarães Pereira, A., Chiaramello, M. and Curvelo, P. (2012) Do-it-yourself Justice - Considerations of Social Media use in a Crisis Situation: The Case of the 2011 Vancouver Riots. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)* (ASONAM '12). IEEE Computer Society, Washington, DC, USA, 720-721. DOI=http://dx.doi.org/10.1109/ASONAM.2012.128

Robins, G., Pattison, P., Woolcock, J., (2004). Missing data in networks: exponential random graph (p∗ ) models for networks with non-respondents. Social Networks 26, 257–283

Sadikov, E., Medina, M., Leskovec, J., Garcia-Molina, H. (2011) Correcting for Missing Data in Information Cascades. ACM International Conference on Web Search and Data Mining (WSDM).

Sakaki, T. Okazaki, M. & Matsuo, Y. (2010) Earthquake shakes Twitter users: real-time event detection by social sensors. Proc. of the 19th Int. Conf. on WWW. ACM.

Tapia, A. H., LaLone, N. J. & Kim, H.-W. (2014). Run amok: Group crowd participation in identifying the bomb and bomber from the Boston marathon bombing.. In S. R. Hiltz, L. Plotnick, M. Pfaf & P. C. Shih (eds.), *ISCRAM*, : ISCRAM Association. ISBN: 978-0-692-21194-6

Tiell, S., Ahuja, A., Lynch, H., Sypics, M. (2016). Ethical algorithms for “sense and respond” systems. Accenture Labs.

Watson, H., Finn, RL. (2013). Privacy and ethical implications of the use of social media during a volcanic eruption: some initial thoughts. Proceedings of the 10th International ISCRAM Conference, 416-420.

Watson, H., Baruh, L., Finn, RL., Scifo, S. (2014) "Citizen (In)Security?: Social Media & Crisis Response". In S. R. Hiltz, L. Plotnick, M. Pfaf & P. C. Shih (eds.), ISCRAM: ISCRAM Association

Yang, D., Zhang, D., Frank, K., Robertson, P., Jennings, E., Roddy, M., Lichtenstern, M. (2014). Providing real-time assistance in disaster relief by leveraging crowdsourcing power. Personal and Ubiquitous Computing