Leveraging Machine Learning for Fake News Detection

Elio Masciari, Vincenzo Moscato, Antonio Picariello and Giancarlo Sperlì
University Federico II, Naples, Italy

Keywords: Fake News Detection, Machine Learning.

Abstract: The uncontrolled growth of fake news creation and dissemination we observed in recent years causes continuous threats to democracy, justice, and public trust. This problem has significantly driven the effort of both academia and industries for developing more accurate fake news detection strategies. Early detection of fake news is crucial, however the availability of information about news propagation is limited. Moreover, it has been shown that people tend to believe more fake news due to their features (Vosoughi et al., 2018). In this paper, we present our complete framework for fake news detection and we discuss in detail a solution based on machine learning. Our experiments conducted on two well-known and widely used real-world datasets suggest that our settings can outperform the state-of-the-art approaches and allows fake news accurate detection, even in the case of limited content information.

1 INTRODUCTION

Social media are nowadays the main medium for large-scale information sharing and communication and they can be considered the main drivers of the Big Data revolution we observed in recent years(Agrawal et al., 2012). Unfortunately, due to malicious user having fraudulent goals fake news on social media are growing quickly both in volume and their potential influence thus leading to very negative social effects. In this respect, identifying and moderating fake news is a quite challenging problem. Indeed, fighting fake news in order to stem their extremely negative effects on individuals and society is crucial in many real life scenarios. Therefore, fake news detection on social media has recently become an hot research topic both for academia and industry.

Fake news detection dates back long time ago(Zhou et al., 2019a) as journalist and scientists always fought against misinformation during human history. Unfortunately, the pervasive use of internet for communication allows for a quicker and wider spread of false information. Indeed, the term *fake news* has grown in popularity in recent years, especially after the 2016 United States elections but there is still no standard definition of fake news (Shu et al., 2017a).

Aside the definition that can be found in literature, one of the most well accepted definition of fake news is the following: *Fake news is a news article that is intentionally and verifiable false and could mislead readers* (Allcott and Gentzkow, 2017). There are two key features of this definition: authenticity and intent. First, fake news includes false information that can be verified as such. Second, fake news is created with dishonest intention to mislead consumers(Shu et al., 2017b).

The content of fake news exhibits heterogeneous topics, styles and media platforms, it aims to mystify truth by diverse linguistic styles while insulting true news. Fake news are generally related to newly emerging, time-critical events, which may not have been properly verified by existing knowledge bases due to the lack of confirmed evidence or claims. Thus, fake news detection on social media poses peculiar challenges due to the inherent nature of social networks that requires both the analysis of their content (Potthast et al., 2017; Guo et al., 2019; Masood and Aker, 2018; Masciari, 2012) and their social context(Shu et al., 2019; Cassavia et al., 2017; Masciari, 2012).

Indeed, as mentioned above fake news are written on purpose to deceive readers to believe false information. For this reason, it is quite difficult to detect a fake news analysing only the news content(Shabani and Sokhn, 2018). Therefore, we should take into account auxiliary information, such as user social engagement on social media to improve the detection accuracy. Unfortunately, the usage of auxiliary information is a non-trivial task as users social engage-

Masciari, E., Moscato, V., Picariello, A. and Sperlì, G.
Leveraging Machine Learning for Fake News Detection.
DOI: 10.5220/0009767401510157
In *Proceedings of the 9th International Conference on Data Science, Technology and Applications (DATA 2020)*, pages 151-157
ISBN: 978-989-758-440-4

ments with fake news produce data that are big, noisy, unstructured and incomplete.

Moreover, the diffusion models for fake news changed deeply in recent years. Indeed, as mentioned above, some decades ago, the only medium for information spreading were newspapers and radio/television but recently, the phenomenon of fakes news generation and diffusion take advantage of the internet pervasive diffusion and in particular of social media quick pick approach to news spreading. More in detail, user consumption behaviours have been affected by the inherent nature of these social media platforms: 1) They are more pervasive and less expensive when compared to traditional news media, such as television or newspapers; 2) It is easier to share, comment on, and discuss the news with friends and followers on social media by overcoming geographical and social barrier.

Despite the above mentioned advantages of social media news sharing, there are many drawbacks, requiring also an suitable pre-processing phase(Mezzanzanica et al., 2015; Boselli et al., 2018). First of all, the quality of news on social media is lower than traditional news organizations due to the lower control of information sources. Moreover, since it is cheaper to provide news online and much faster and easier to spread through social media, larger volumes of fake news are produced online for a variety of purposes, such as political gain and unfair competition to cite a few.

Fake News Examples. Some well known examples of fake news across history are mentioned below: a) During the second and third centuries AD, false rumours were spread about Christians claiming that they engaged in ritual cannibalism and incest[1]; b) In 1835 The New York Sun published articles about a real-life astronomer and a made-up colleague who, according to the hoax, had observed bizarre life on the moon[2]; c) More recently we can cite some news like, Paul Horner, was behind the widespread hoax that he was the graffiti artist Banksy and had been arrested; a man has been honored for stopping a robbery in a diner by quoting Pulp Fiction; and finally the great impact of fake news on the 2016 U.S. presidential election, according to CBS News[3].

Furthermore, in 2018 BuzzFeed News compiled a list of 50 most viral false stories on Facebook and measured their total engagement on the platform. And in spite of a prediction from Facebook's top anti-misinformation product manager that these articles would see a decline in engagement in 2018, the top-performing hoaxes generated roughly 22 million total shares, reactions, and comments on Facebook between Jan. 1 and Dec. 9, 2018.

Psychological Aspects behind Fake News. The influential power of fake news has been explained by several psychological theories. Fake news mainly targets people by exploiting their vulnerabilities. There are two major factors which make consumers naturally vulnerable to fake news (Shu et al., 2017a; Zhou et al., 2019b): 1) *Naive Realism* as people tend to believe that their perceptions of reality are the only accurate views, while others who disagree are regarded as uninformed or irrational and 2) *Confirmation Bias* as people prefer to receive information that confirms their beliefs.

Moreover, people are more susceptible to certain kinds of (fake) news due to to the way newsfeed appears on their homepages in social media thus amplifying the psychological challenges to dispelling fake news. Indeed, people trust fake news owing to two psychological factors(Shu et al., 2017a): 1) *social credibility*, which means people are more likely to recognize a source as trustworthy if others recognize the source as reliable, mainly when there is not enough information to assess the truthfulness of the source; 2) *frequency heuristic*, which means that people may obviously favour information they hear repeatedly, although it is fake news.

Our Approach in a Nutshell. Fake news detection problem can be formalized as a classification task thus requiring features extraction and model construction sub-taks. The detection phase is a crucial task as it is devoted to guarantee users to receive authentic information. We will focus on finding clues from news contents.

Our goal is to improve the existing approaches defined so far when fake news is intentionally written to mislead users by mimicking true news. More in detail, traditional approaches are based on verification by human editors and expert journalists but do not scale to the volume of news content that is generated in online social networks. As a matter of fact, the huge amount of data to be analyzed calls for the development of new computational techniques. It is worth noticing that, such computational techniques, even if the news is detected as fake, require some sort of expert verification before being blocked. In our framework, we perform an accurate pre-processing of news data and then we apply several approaches for analyzing text and multimedia contents. The approach we discuss in detail in this paper is based on machine

[1]https://en.wikipedia.org/wiki/Fake news

[2]http://www.snelgraphix.net/the-snelgraphix-designing-minds-blog/tag/google+I%E2%80%99m+feeling+stellar

[3]https://www.businessinsider.com/banksy-arrest-hoax-2013-2

learning techniques. In this respect, we implemented several algorithms and we compared them as will be better explained in the experimental section in order to find out the most suitable one for the fake news scenario.

2 OUR FAKE NEWS DETECTION FRAMEWORK

Our framework is based on news flow processing and data management in a pre-processing block which performs filtering and aggregation operation over the news content. Moreover, filtered data are processed by two independent blocks: the first one performs natural language processing over data while the second one performs a multimedia analysis.

The overall process we execute for fake news detection is depicted in Figure1.

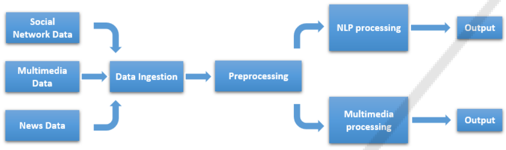

Figure 1: The overall process at a glance.

In the following we describe each module in more detail.

Data Ingestion Module. This module take care of data collection tasks. Data can be highly heterogeneous: social network data, multimedia data and news data. We collect the news text and eventual related contents and images.

Pre-processing Module. This component is devoted to the acquisition of the incoming data flow. It performs filtering, data aggregation, data cleaning and enrichment operations.

NLP Processing Module. It performs the crucial task of generating a binary classification of the news articles, i.e., whether they are fake or reliable news. It is split in two submodules. The *Machine Learning* module performs classification using an ad-hoc implemented Logistic Regression algorithm (the rationale for this choice will be explained in the experimental section) after an extensive process of feature extraction and selection TF-IDF based in order to reduce the number of extracted features. The *Deep Learning* module classify data using Google Bert algorithm after a tuning phase on the vocabulary. It also perform a binary transformation and eventual text padding in order to better analyze the input data.

Multimedia Processing Module. This module is tailored for Fake Image Classification through Deep Learning algorithms, using ELA (Error Level Analysis) and CNN.

Due to space limitation, we discuss in the following only the details of the deep learning module and the obtained results.

2.1 The Software Architecture

The software implementation of the framework described above is shown in Figure 2.

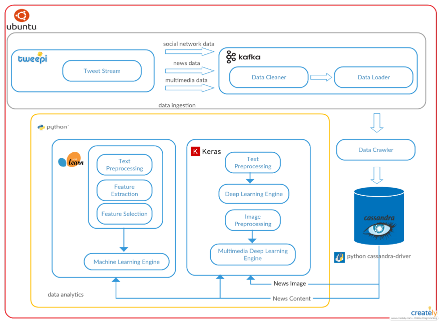

Figure 2: Our fake news detection framework.

Herein: the data ingestion block is implemented by using several tools. As an example for Twitter data we leverage Tweepy[4], a Python library to access the Twitter API. All tweets are downloaded through this library. Filtering and aggregation is performed using Apache Kafka[5] which is able to build real-time data pipelines and streaming apps. It is scalable, fault-tolerant and fast thus making our prototype well-suited for huge amount of data.

The data crawler uses the Newspaper Python library[6] whose purpose is extracting and curating articles. The analytical data archive stores pre-processed data that are used for issuing queries by traditional analytical tools. We leverage Apache Cassandra[7] as datastore because it provides high scalability, high availability, fast writing, fault-tolerance on commodity hardware or cloud infrastructure. The data analytics block retrieves news contents and news images from Cassandra DB that are pre-processed by the Machine Learning module using Scikit Learn library[8] and by Deep Learning module using Keras library[9]. Image content is processed by the Multimedia Deep Learning module using Keras library.

In the following we will briefly describe how the overall process is executed. Requests to the Cassandra

[4]https://www.tweepy.org/
[5]https://kafka.apache.org/
[6]https://newspaper.readthedocs.io/en/latest/
[7]http://cassandra.apache.org/
[8]https://scikit-learn.org/stable/
[9]https://keras.io/

DB are made through remote access. Each column in Cassandra refers to a specific topic and contains all news belonging to that topic. Among all news, those having a valid *external link* value are selected. In this way, the news content can be easily crawled. As the link for each news is obtained, a check is performed in order to verify the current state of the website. If the website is still running, we perform the article scraping. The algorithm works by downloading and parsing the news article, then, for each article, title, text, authors, top image link, news link data are extracted and saved as a JSON file in Cassandra DB.

Finally, three independent analysis are then performed by three ad-hoc Python modules we implemented. The first two perform text classification, and the last one images classification. Concerning the text analysis, the problem being solved is a binary classification one where class 0 refers to reliable news and class 1 refers to fake ones.

2.1.1 The Machine Learning Module

As mentioned above the goal of the Machine Learning Module is to produce a binary classification on a text dataset. Thus, a news article will be labelled as 0 if it is recognised as *Real*, and as 1 if it is recognised as *Fake*. We devise a supervised approach since the dataset we worked on is fully labelled. The Machine Learning implementation has been chosen by comparing most of the available classifiers provided by the Scikit- Learn library.

It has been developed using a Python 3 kernel in Jupyter, that is a web-based interactive development environment for code, and data. It is flexible, extensible, modular and configurable to support a wide range of workflows in data science, scientific computing, and machine learning. We choose Python as it is interactive, interpreted, modular, dynamic, object-oriented, portable and extensible thus offering an high flexibility for our purposes.

More in detail, the following libraries have been used: i) Scikit- Learn: a simple and efficient tool for data mining and data analysis; ii) Spacy: an open-source software library for advanced Natural Language Processing, iii) Numpy: a library for Python programming language that offer support for large, multi- dimensional arrays and matrices, along with a large collection of high level mathematical functions to operate on these arrays, iv) Pandas: an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming and v) Matplotlib: a plotting library for the Python programming language and its numerical mathematics extension NumPy.

In order to choose the most suitable classification method, we performed an extensive tuning phase by comparing several algorithms. The performances have been evaluated on several test sets by comparing several accuracy measure like Accuracy, Precision, Recall, F1 measure, Area Under Curve (AUC) (Flach and Kull, 2015) (reported in Figure 3) and execution times (reported in Figure 4). The results are shown for LIAR datasets but the same trend has been observed for each data set being analyzed.

As it is easy to see, the best model in terms of accuracy turns out to be Logistic Regression, so we decided to perform a parameter optimization for this algorithm as it exhibits the best results on each efficiency and effectiveness measure. It is worth noticing that, classifiers based on tree construction executes much slower because of the training step.

We briefly recall here, that logistic regression instead is a statistical model that leverages the logit function to model a binary dependent variable [13], i.e., a linear combination of the observed features:

$$log \frac{p}{1-p} = \beta_0 + \beta_1 \dot{x}$$

Logistic Regression outputs the probabilities of a specific class that are then used for class predictions. The logistic function exhibits two interesting properties for our purposes: 1) it has a regular “s” shape; 2) Its output is bounded between 0 and 1.

Compared with other models, Logistic Regression offers the following advantages: 1) it is easily interpretable; 2) Model training and prediction steps are quite fast; 3) Only few parameters has to be tuned (the regularization parameter); 4) It performs well even with small datasets; 5) It outputs well-calibrated predicted probabilities. Nevertheless, there are some drawbacks as the need of a linear relationship between the features and the log-odds of the response and it is not able to automatically learn feature interactions In order to tune the algorithm we leveraged the functionalities offered by SciKit.

3 OUR BENCHMARK

In this section we will describe the fake news detection process and the datasets we used as a benchmark for our algorithms.

3.1 Dataset Description

Liar Dataset. This dataset includes 12.8K human labelled short statements from fact-checking website

Classifier	Accuracy	Precision	Recall	F1	TP	FP	TN	FN	AUC
SGD	0.620	0.627	0.738	0.678	1020	605	564	362	0.610
Naïve Bayes	0.604	0.591	0.871	0.704	1205	833	336	177	0.580
Linear SVC	0.600	0.623	0.662	0.642	915	553	616	467	0.595
Random Forest	0.586	0.588	0.790	0.674	1093	765	404	289	0.568
Logistic Regression	0.627	0.631	0.758	0.684	1034	603	566	348	0.616
Nearest Neighbor	0.563	0.588	0.646	0.616	894	625	544	488	0.556
Decision Tree	0.562	0.592	0.616	0.604	852	585	584	530	0.558
Gradient Boost	0.603	0.592	0.860	0.701	1189	818	351	193	0.580
Perceptron	0.578	0.572	0.749	0.638	1088	770	399	294	0.571
Passive Aggressive	0.593	0.599	0.768	0.702	935	533	586	497	0.591

Figure 3: Classifier Effectiveness Comparison.

Classifier	Training Time	Test Time
SGD	0.06	0.03
Naïve Bayes	0.02	0.004
Linear SVC	0.27	0.002
Random Forest	3.57	0.06
Logistic Regression	0.22	0.002
Nearest Neighbor	0.019	4.17
Decision Tree	9.18	0.009
Gradient Boost	22.7	0.02
Perceptron	0.08	0.05
Passive Aggressive	0.09	0.04

Figure 4: Execution Times comparison.

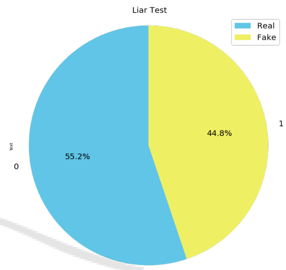

Figure 5: LIAR Test Dataset in a short.

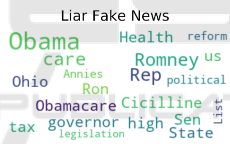

Figure 6: Liar Fake Wordclouds.

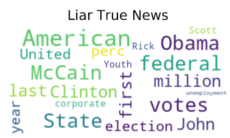

Figure 7: Liar Real Wordclouds.

Politifact.com. Each statement is evaluated by a Politifact.com editor for its truthfulness. The dataset has six fine-grained labels: pants-fire, false, barely-true, half-true, mostly-true, and true. The distribution of labels is relatively well- balanced. For our purposes the six fine-grained labels of the dataset have been collapsed in a binary classification, i.e., label 1 for fake news and label 0 for reliable ones. This choice has been made due to binary Fake News Dataset feature. The dataset is partitioned into three files: 1) Training Set: 5770 real news and 4497 fake news; 2) Test Set: 1382 real news and 1169 fake news; 3) Validation Set: 1382 real news and 1169 fake news. In Figure5 we show the distribution of real and fake news for the test dataset.

The three subsets are well balanced so there is no need to perform oversampling or undersampling. The corresponding Wordclouds for fake news is reported in Figure 6. It is easy to see that news are mainly related to United States. Fake news topics are collected about Obama, Obamacare, Cicilline, Romney.

On the other side real news topics depicted in Figure 7 refer to McCain, elections and Obama.

The processed dataset has been uploaded in Google Drive and, then, loaded in Colab's Jupyter as a Pandas Dataframe. It has been added a new column with the number of words for each row article. By this column it is possible to obtain the following statistical information: count 15389.000000, mean 17.962311, std 8.569879, min 1.000000, 25% 12.000000, 50% 17.000000, 75% 22.000000, max 66.000000. These statistics show that there are articles with only one word in the dataset, so it has been decided to remove all rows with less than 10 words as they are considered poorly informative. The resulting dataset contains 1657 less rows than the original one. The updated statistics are reported in what follows: count 13732.000000, mean 19.228663, std 8.192268, min 10.000000, 25% 14.000000, 50% 18.000000, 75% 23.000000, max 66.000000. Finally, the average number of words per article is 19.

FakeNewsNet. This dataset has been built by gathering information from two fact-checking websites to obtain news contents for fake news and real news such as PolitiFact and GossipCop. In PolitiFact, journalists and domain experts review the political news and provide fact-checking evaluation results to claim news articles as fake or real. Instead, in GossipCop, entertainment stories, from various media outlets, are evaluated by a rating score from on the scale of 0 to 10 as the degree from fake to real. The dataset contains 900 political news and 20k gossip news and has only two labels: true and false.

This dataset is publicly available by the functions provided by the FakeNewsNet team and the Twitter API. As mentioned above, FakeNewsNet can be split in two subsets: GossipCop and Politifact.com. We decided to analyse only political news as they produce worse consequences in real world than gossip ones. The dataset is well balanced and contains 434 real news and 367 fake news. Most of the news regards the US as it has already been noticed in LIAR. Fake news topics concern Obama, police, Clinton and Trump while real news topics refer to Trump, Republicans and Obama. Such as the LIAR dataset, it has been added a new column and the following statistical information have been obtained: count 801, mean 1459.217228, std 3141.157565, min 3, 25% 114, 50% 351, 75% 893, max 17377.

The average number of words per articles in Politifact dataset is 1459, which is far longer than the average sentence length in Liar Dataset that is 19 words per articles. Such a statistics confirmed our belief that it would be better to compare the model performances on datasets with such different features.

4 EVALUATION

In order to show that the model we implemented outperforms the results of the current approaches, we preliminary report in Figure 8 the best results obtained for the other approaches commonly used in literature for LIAR datasets and in Figure 9 the results we obtained for the logistic regression algorithm we implemented that gave us the best results in the fake news scenario.

We compared the performances on well-established evaluation measure like: Accuracy, Precision, Recall, F1 measure, Area Under Curve (AUC) (Flach and Kull, 2015) and the values reported in the obtained confusion matrices for each algorithm, i.e., True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN).

		Liar			
Model	**Feature**	Acc	Pre	Rec	F1-Score
SVM	Lexical	.56	.56	.56	.48
SVM	Lexical +Sentiment	.56	.57	.56	.48
LR	Lexical +Sentiment	0.56	.56	.56	.51
Decision Tree	Lexical +Sentiment	.51	.51	.51	.51
Adaboost	Lexical +Sentiment	.56	.56	.56	.54
Naive Bayes	Unigram (TF-IDF)	.60	.60	.60	.57
Naive Bayes	Bigram (TF-IDF)	**.60**	**.59**	**.60**	**.59**
k-NN	Empath Features	.53	.53	.53	.53

Figure 8: Comparison against state of the art approaches on LIAR datataset.

ACC	PREC	REC	F1	AUC	TN	TP	FN	FP
0.635	0.629	0.796	0.702	0.621	520	1101	281	649

Figure 9: Our results on LIAR datataset.

We hypothesize that our results are quite better due to the fine feature selection we performed, a better pre-processing step and the proper text transformation and loading.

ACC	PREC	REC	F1	AUC	TN	TP	FN	FP
0.635	0.629	0.796	0.702	0.621	520	1101	281	649

Figure 10: Our results on Polifact datataset.

In Figure 10 we report the results we obtained on Polifact dataset.

For the sake of completeness, we report in Figure 11 and Figure 12 the detailed confusion matrices obtained for LIAR and Polifact datasets.

5 CONCLUSION AND FUTURE WORK

In this paper, we investigated the problem of fake news detection by machine learning algorithms. We developed a framework the leverage several algorithms for analyzing real-life datasets and the results we obtained are quite encouraging. In particular, we

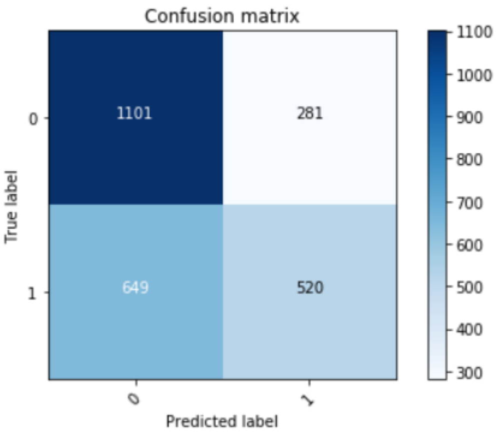

Figure 11: Confusion Matrix for LIAR dataset.

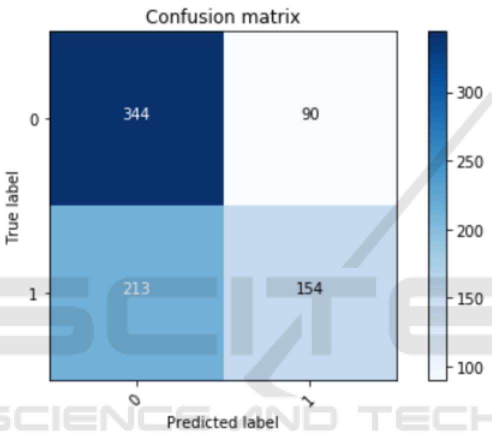

Figure 12: Confusion Matrix for Polifact dataset.

found that the most accurate results can be obtained with logistic regression based algorithms. As a future work, we would like to extend our analysis by better considering also user profiles' features and some kind of dynamic analysis of news diffusion mechanism in our fake news detection model.

REFERENCES

Agrawal et al., D. (2012). Challenges and opportunities with big data. A community white paper developed by leading researchers across the United States. Technical report, Purdue University.

Allcott, H. and Gentzkow, M. (2017). Social media and fake news in the 2016 election. Working Paper 23089, National Bureau of Economic Research.

Boselli, R., Cesarini, M., Mercorio, F., and Mezzanzanica, M. (2018). Classifying online job advertisements through machine learning. *Future Generation Computer Systems*, 86:319 – 328.

Cassavia, N., Masciari, E., Pulice, C., and Saccà, D. (2017). Discovering user behavioral features to enhance information search on big data. *TiiS*, 7(2):7:1–7:33.

Culpepper, J. S., Moffat, A., Bennett, P. N., and Lerman, K., editors (2019). *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM 2019, Melbourne, VIC, Australia, February 11-15, 2019*. ACM.

Flach, P. A. and Kull, M. (2015). Precision-recall-gain curves: PR analysis done right. In Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 838–846.

Guo, C., Cao, J., Zhang, X., Shu, K., and Yu, M. (2019). Exploiting emotions for fake news detection on social media. *CoRR*, abs/1903.01728.

Masciari, E. (2012). SMART: stream monitoring enterprise activities by RFID tags. *Inf. Sci.*, 195:25–44.

Masood, R. and Aker, A. (2018). The fake news challenge: Stance detection using traditional machine learning approaches.

Mezzanzanica, M., Boselli, R., Cesarini, M., and Mercorio, F. (2015). A model-based approach for developing data cleansing solutions. *Journal of Data and Information Quality (JDIQ)*, 5(4):1–28.

Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., and Stein, B. (2017). A stylometric inquiry into hyperpartisan and fake news. *CoRR*, abs/1702.05638.

Shabani, S. and Sokhn, M. (2018). Hybrid machine-crowd approach for fake news detection. In *2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)*, pages 299–306. IEEE.

Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H. (2017a). Fake news detection on social media: A data mining perspective. *CoRR*, abs/1708.01967.

Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H. (2017b). Fake news detection on social media: A data mining perspective. *CoRR*, abs/1708.01967.

Shu, K., Wang, S., and Liu, H. (2019). Beyond news contents: The role of social context for fake news detection. In (Culpepper et al., 2019), pages 312–320.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380):1146–1151.

Zhou, X., Zafarani, R., Shu, K., and Liu, H. (2019a). Fake news: Fundamental theories, detection strategies and challenges. In (Culpepper et al., 2019), pages 836–837.

Zhou, Z., Guan, H., Bhat, M. M., and Hsu, J. (2019b). Fake news detection via nlp is vulnerable to adversarial attacks. *arXiv preprint arXiv:1901.09657*.