Master 2 Research Project

Report

Detecting Misinformation and its Sources on Social Media

by

Elie Alex KAMENI NGANGUE

Master 2 Research DSML

ETIS UMR8051, CY Cergy Paris University / ENSEA / CNRS
6 avenue du Ponceau, 95014 Cergy-Pontoise Cedex, France

Defended on 30 March 2022

In front of the jury composed of:

Pr. Son VU	ETIS/CYU/ENSEA/CNRS	President
Pr. Dimitris KOTZINOS	ETIS/CYU/ENSEA/CNRS	Supervisor
M. Malek	ENSEA	Professor advisor

Detecting Misinformation and its Sources on Social Media

Elie Alex KAMENI NGANGUE
University of CY Cergy Paris
ENSEA
elie-alex.kameni-ngangue@ensea.fr

***Abstract*—Over the last decade, social networks have become increasingly popular for news consumption due to their easy access, rapid dissemination, and low cost. However, social networks also allow for the wide spread of ”fake news”, i.e., the spread either by accident, lack of proper knowledge or deliberately of news having false information. Fake news on social networks can have significant negative effects on society. Therefore, detecting fake news on social networks has recently become an emerging research area that is receiving considerable attention. There are several types of spreaders, and we focus on those who spread misinformation in social networks, and our research particularly focuses on those who spread it on Twitter during the Covid-19 health crisis. In this work we propose a solution that can be used by users to detect and filter sites having false and misleading information on the one hand, and by social network administrators to reduce the spread of false information on their platform on the other. We use simple and carefully selected characteristics of the title of the site where the information comes from and the content of the tweet to accurately identify false information. The approach used in this paper is to adopt a lightweight and efficient architecture for misinformation detection in the first step and a recommender-based architecture to determine which users are more likely to share the misinformation detected. Our Bidirectional GRU model performed 95.3% f1-score on the COVID-19 health misinformation dataset thus placing itself above the state of the art on this data set and our top-10 tweets not to recommend to a user model scored around 30% [1].**

I. Introduction

Social media has been used to gather information about large-scale events, such as fires, earthquakes, and other disasters. Social media provides users with a large-scale, easy-to-use platform that is not possible with traditional media. Understanding the spread of information in social media provides additional context, such as knowing where information originates and how it changes as it moves from one user to another until the end of its lifecycle. The normal social media user applies this knowledge to assess the reliability and accuracy of this information.

Throughout the social media experience, users face the problem of determining the authenticity and quality of data. It is difficult to assess the trustworthiness of a source on such a user-generated content platform, where information sources may mistakenly or intentionally spread false information. This in turn leads to the spread of polluted information.

Social networks thus make each user a self-publisher without editing, without verification of factual accuracy and clearly without responsibility. The use of technology by people to support lies, deception, misdirection, fraud, image control, propaganda has become a reality with online social networks like Facebook and Twitter being used for purposes for which they were not intended. Twitter has emerged as one of the most popular microblogging sites. The lack of accountability and verifiability provides users with an excellent opportunity to spread specific ideas on the network.

All major social media are affected by ”fake news”. And the world’s most visited, YouTube, too. According to a study conducted by researchers from the University of Ottawa in Canada on a sample of videos, more than a quarter of the most viewed videos on YouTube contain misleading information about Covid-19 disease [1]. Nevertheless, measures have been taken to limit the spread of misinformation.

In September 2020, YouTube launched in France, after other countries, the fact-checking of videos [2]. In December 2020, TikTok announced the implementation of measures to fight misinformation about vaccines. The social network announced to strengthen its moderation work and rules to be able to remove false information about vaccines [3]. Until now, Facebook had been content to reduce the visibility of these publications. Facebook announced on Thursday, December 2, 2021, that it plans to remove ”false claims about Covid-19 vaccines that have been disproven by public health experts on Facebook and Instagram” in the coming weeks.

There are methods using machine learning and natural language processing (NLP) techniques to automate the process. In this article, we present the approach we used for detecting misinformation and finding which users are most likely to share the information. The rest of this paper is subdivided as follows: first, we start by presenting the literature, then the different datasets on which we conducted our experiments. After that, we formulate the problem addressed in this paper and present the different architectures used to solve the problem. Finally, we present our results.

II. Related works

A. Categorization of misinformation detection methods

Misinformation detection is defined as an observation that deviates greatly from other observations and thereby arouses suspicion that it was generated by a different mechanism [4]. It seems to be a classification problem, which has the same setting as text categorization. Traditional text categorization tasks,

[1] code repository of this paper

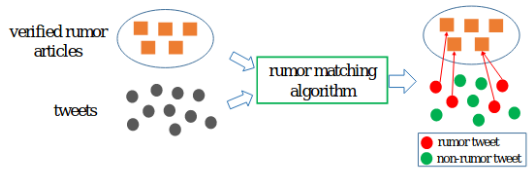


Fig. 1: Rumour detection as a text matching task [6]

where the content is mostly organic and written/compiled to be distinguishable, e.g., sports news articles are meant to be different from political news. By contrast, misinformation posts are deliberately made seemingly real and accurate.

Based on the information that a method mainly utilizes, Liang et al. [5] proposed a categorization of detection methods as follows : content-based misinformation detection, context-based misinformation detection, propagation-based misinformation detection, early detection of misinformation.

1) Content-based misinformation detection: In content-based misinformation detection, misinformation is directly detecting based on its content, such as text, images and video. For example, some studies focus on retrieving all posts related to a known piece of misinformation [6, 7]. This stream of research is more of a text matching problem, where the targeted posts are those very similar or duplicate ones of an original misinformation post (Figure 1). In order to extend the limits of text matching methods, supervised learning methods have been studied to identify misinformation [8, 9]. However, post-based methods can be overly sensitive to misinformation content.

2) Context-based misinformation detection: In this approach, misinformation is detected based on the contextual information available in social media, such as locations and time. This contextual information is usually jointly used with other information to ease the detection, or directly vectorized and being used as added features [10, 11]

3) Propagation-based misinformation detection: In this approach, misinformation is detected based on the propagation patterns, i.e., how misinformation circulates among users. Since intentional spreaders of misinformation may manipulate the content to make it seem very real, it is particularly challenging to obtain useful features from content for these emerging applications. To address this problem, some recent work concentrates on modelling the propagation of messages in a social network. A key intuition of using propagation information is that the direction of information exchange can reveal community structures and personal characteristics [12]

B. Deep learning for misinformation detection

Although many techniques are being used to detect misinformation in social network data, deep learning is one of the better approaches to use [4]. However, the same type of misinformation problems has been solved with various deep learning techniques classified into three groups types of deep learning techniques which are dependent on different data characteristics and used to automatically identify misinformation. (Figure 2).

1) Discriminative model for detecting misinformation: A variety of discriminative models used social content and context-based features for misinformation detection. In recent years, to tackle the problem of misinformation, several studies have been conducted and revealed some promising preliminary results.

Convolutional Neural Network (CNN): CNN is one of the most popular and widely used models for the state-of-the-art of many computer vision tasks. However, recently, it has been extensively applied in the NLP community as well (Jacovi et al. [13]). For example, Chen et al. (2017) [14] introduced a convolutional neural network-based classification method with single and multi-word embedding for identifying both rumour and stance tweets. First, they used word embedding to convert each word in the tweet into a vector with randomly initialize of the word embedding matrix. Then, they learned the embedding weights during the training process. Second, they have concatenated these word vectors to produce a matrix representing the sentence.

Recurrent Neural Network (RNN): RNN uses the sequential information in the network which is essential in many applications where the embedded structure in the data sequence conveys useful knowledge (Alkhodair et al. 2020 [15]). The advantage of RNN is its ability to better capture contextual information. To detect misinformation, existing methods rely on handcrafted features to employ machine learning algorithms that require a huge manual effort. To guard against this issue, the earliest adoption of RNNs for rumour detection is reported in Ma et al. (2016) [16].

2) Generative model for detecting misinformation: Several existing works for misinformation detection are based on syntactic and lexical patterns or features of opinion. However, this information is not always present in the dataset we use. Generative models are therefore used to overcome this limitation

Variational Auto Encoder (VAE): VAE models make strong assumptions concerning the distribution of latent variables. The use of a variational approach for latent representation learning results in an additional loss component [4]. Qian et al. (2018) [17] proposed a generative conditional VAE model to extract new patterns by analyzing a user's past meaningful responses on true and false news articles and played a vital role in detecting misinformation on social media.

Generative Adversarial Network (GAN): Given a training set, this technique learns to generate new data with the same statistics as the training set. Ma et al. (2019) [18] proposed a generative adversarial network model to make automated misinformation detection more robust and efficient and is designed to identify powerful features related to uncertain or conflicting voice production and rumours.

C. Misinformation spreads

The openness and timeliness of social media have to a great extent encouraged the creation and spread of misinformation,

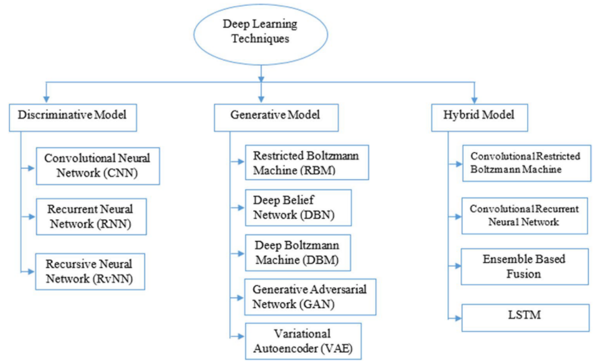


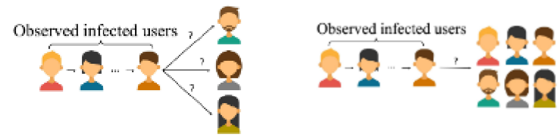
Fig. 2: Classification of deep learning models [4]

such as rumour, spam, and fake news. Misinformation can affect every aspect of life such as the social, political, economic, stock market, emergency response during natural disasters, and crisis events. It aims to intentionally or unintentionally mislead public opinions, influence political elections, and threaten public security and social stability (Wu et al. [5]). Nowadays, it has become easier to spread misinformation quickly due to social network platforms such as Facebook, Twitter, and Sina Weibo. When people engage in conversation, one can share information that is stated to be factual, but that may not always be true. Additionally, fraudulent users share misleading information to look for personal gain in some way.

The prediction of information diffusion, also known as cascade prediction, has been studied over a wide range of applications, such as product adoption [19, 20], epidemiology [21], social networks [22] and the spread of news and opinions [23]. Recent works [24, 25, 26] on diffusion prediction took advantage of the success of deep learning techniques by modelling information diffusion as sequential data based on recurrent neural networks (RNNs) and achieved promising performances. Existing works [27, 28] also explored the social graph information which is available when information diffusion spreads through a social network service for diffusion prediction.

However, as shown in Figure 3 [29], previous works focused on either microscopic diffusion prediction which aims at guessing the next infected user or macroscopic diffusion prediction which estimates the total numbers of infected users during the diffusion process. A unified model can use more information in the training data especially for macroscopic diffusion prediction, e.g. previous works [24, 25] considered cascade size prediction as a regression problem and discarded the information of detailed infected users and the ordering of their infections.

These works do not distinguish the type of information spread, therefore consider that a misinformation is shared in the same way as a truth information. In this paper, we are interested only in the diffusion of misinformation in a network, and we propose a model based on the history of tweets shared by users to predict the next ones that will be of interest and that he could share.



(a) Who's the next infected user?

(b) How many infected users at last

Fig. 3: Illustrative examples for microscopic next infected user prediction (left) and macroscopic cascade size prediction (right) [29]. Conventionally, we usually use "infected" to indicate that a user is "influenced" by an information item

TABLE I: Examples of real and fake news from the dataset. Fake news is collected from various sources. Real news is collected from verified Twitter accounts. [30]

Label	Source	Text
Fake	Twitter	#Watch Italian Billionaire commits suicide by throwing himself from 20th Floor of his tower after his entire family was wiped out by #Coronavirus #Suicide has never been the way, may soul rest in peace May God deliver us all from this time
Fake	Twitter	It's being reported that NC DHHS is telling hospitals that if they decide to do elective surgeries, they won't be eligible to receive PPE from the state. The heavy hand of government. I hope Secretary Cohen will reverse course. #NCDHHS #COVID19NC #ncpol
Real	Twitter (WHO)	Almost 200 vaccines for #COVID19 are currently in clinical and pre-clinical testing. The history of vaccine development tells us that some will fail and some will succeed-@DrTedros #UNGA #UN75
Real	Twitter (CDC)	Heart conditions like myocarditis are associated with some cases of #COVID19. Severe cardiac damage is rare but has occurred even in young healthy people. CDC is working to understand how COVID-19 affects the heart and other organs.

III. Dataset description

A. *Misinformation detection dataset*

The dataset used to train our misinformation detection model is a dataset of real and fake information on COVID-19 [30] :

- Real : Tweets from verified sources and give useful information on COVID-19.
- Fake : Tweets, posts, articles which make claims and speculations about COVID-19 which are verified to be not true.

Table I gives some examples of real and fake news from the dataset

TABLE II: Numeric feature of the dataset

Attribute	Fake	Real	Combined
Unique words	19728	22916	37503
Avg words per post	21.65	31.97	27.05
Avg chars per post	143.26	218.37	182.57

TABLE III: Distribution of data across classes and splits.

Split	Real	Fake	Total
Training	3360	3060	6420
Validation	1120	1020	2140
Test	1120	1020	2140
Total	5600	5100	10700

From Table II, we observe that, in general, real news are longer than fake news in terms of average number of words and characters per post. The vocabulary size (i.e., unique words) of the dataset is 37,505 with 5141 common words in both fake and real news. The original dataset is split into train (60%), validation (20%), test (20%). Table III shows the class-wise distribution of all data splits. The dataset is class-wise balanced as 52.34% of the samples consist of real news and 47.66% of the data consists of fake news. We analyse the dataset on token-level. The 10 most frequent tokens after removing stop words are:

- Fake: coronavirus, covid19, people, will, new, trump, says, video, vaccine, virus.
- Real: covid19, cases, new, tests, number, total, people, reported, confirmed, states.
- Combined: covid19, cases, coronavirus, new, people, tests, number, will, deaths, total.

B. Misinformation spreads dataset

For our study of misinformation dissemination in a social network, we collected 1 million out of 230 million which correspond to 220k conversations, 490k users and 880k replies [31].

Since our use case is to study the spreading of misinformation in a social network, only tweets and replies under false articles were considered. Then we added all the tweets with a tweet under a fake article quoting it. We then filtered out the tweets of low-scattering users (who appear less than 5 times and whose tweets have less than 5 replies) and removed low-engagement conversations (less than 10 tweets). We kept 1210 tweets, 297 users, resulting in a user interaction dataset with the tweets of about 3k rows.

IV. Problems formulation

A. Misinformation detection formulation

The problem of detecting misinformation is closely related to the problem of classifying texts created by a set of users linked to a set of domains into n classes (if n is 2, we talk about binary classification).

Let U be a set of users, $T = \{t_1, ..., t_n\}$ the set of textual data created by them. Each text is t_i is created at a date d_i by a user u_i of U, linked to a context c_i and of class d_i. The problem of detecting misinformation is therefore to find a function f such that $d_i = f(t_i, u_i, ...)$

The search of the function f involves the optimization of a loss function L. Once determined, we evaluate the performance of our dataset on new inputs.

The loss function used here is Cross Entropy defined as:

$$L(p, y) = -(ylog(p) + (1 - y)log(1 - p)) \tag{1}$$

The metrics that allow us to evaluate the performance of our model are f1 score and ROC AUC.

An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters True Positive Rate and False Positive Rate. AUC supplies an aggregate measure of performance across all possible classification thresholds. One way of interpreting AUC is as the probability that the model ranks a random positive example more highly than a random negative example.

F1-score is the weighted average of Precision (ratio of correctly predicted positive observations to the total predicted positive observations) and Recall (the ratio of correctly predicted positive observations to all observations in actual class - yes).

B. Misinformation spreading formulation

Given user set V, cascade set C, each cascade $c_i \in C$ is a sequence of users $[v_i^1, v_i^2, ..., v_i^{|c_i|}]$ ranked by their infection timestamps where $|c_i|$ is the size of the cascade c_i, i.e. the number of users infected by the corresponding item. Also, an underlying social graph $G = (V, E)$ among users will be available when information diffusion occurs on a social network service. The social graph G will be considered as additional structural inputs for diffusion prediction.

Microscopic Diffusion Prediction: aims at predicting the next infected user v_i^{k+1} given previously infected users $\{v_i^1, v_i^2, ..., v_i^k\}$ in cascade c_i for $k = 1, 2..., |c_i| - 1$.

Macroscopic Diffusion Prediction: aims at predicting the eventual size $|c_i|$ of the cascade c_i, i.e. the total number of infected users, given the first K infected users $\{v_i^1, v_i^2, ..., v_i^k\}$.

In this paper, we use the approach based on finding a user representation projection space such that any tweet whose representation is close is one that can be shared by the user.

This method assumes that whether a user shares information or not is based primarily on the content of the information. It is inspired by the virus propagation model.

V. Experiments and results

A. Models

1) Misinformation detection models: We started by reproducing the results obtained in the paper with the SVM (support vectors machine) and logistic regression models. We also used the XGBoost model.

Then, we replaced the Term-frequency layer by transformers encoder layer (all-MiniLM-L6-v2 from Huggingface) to have a better representation of the tweets. It is a sentence-transformer model: it maps sentences and paragraphs to a dense 384-dimensional vector space and can be used for tasks like clustering or semantic search [32]. Other architectures such as BiGRU [33] (Biderectional gated recurrent units), LogREG, NBSVM (NBSVM is a text classification model proposed by Wang and Manning in 2012 that takes a linear

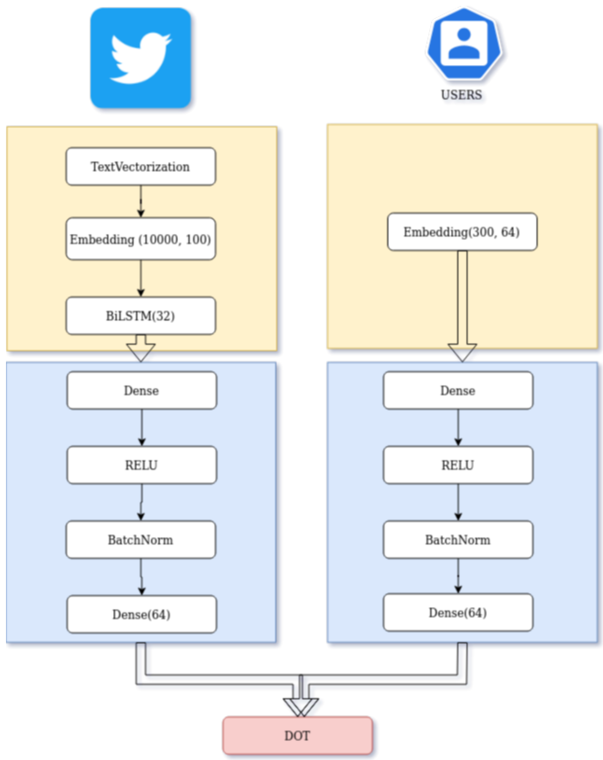


Fig. 4: Misinformation spreads models

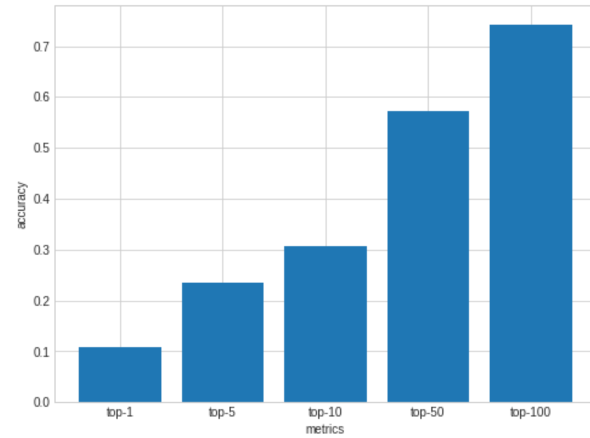


Fig. 5: Misinformation spreads model score

model such as SVM (or logistic regression) and infuses it with Bayesian probabilities by replacing word count features with Naive Bayes log count ratios) and the pre-trained model BERT [34] have been used for a better representation of the current context in order to increase the performance of our misinformation detection model.

2) Misinformation spreads models: The proposed approach is based on collaborative recommendation systems (matrix factorization with the projection parameter as a hyperparameter of our model). Each user and tweet are represented in the system by a tensor of the same dimension. The tweets that are closest to the users are then the ones they are more likely to share. [35]

On the one hand, we have a model for the representation of tweets consisting of an embedding layer followed by a BiLSTM layer: CandidateEmbegging On the other hand, an embedding layer of users having shared the information: QueryEmbedding. These two models are followed by Dense, ReLU and BatchNorm layers in order to have an encoding of queries and candidates on the same space (Figure 4).

B. Experiments settings

All our experiments were conducted on the same machine with an NVIDIA GeForce GTX 1650 GPU, 16 GB RAM, Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz. We work under Linux version 5.4.0-105-generic.

We use sklearn [36] for the SVM and LR implementations, XGBoost [37] for the XGBoostClassifier, sentence-transformer [38] for the MiniLM sentence encoding model and the Ktrain [39] library for the fast development of our different deep models. We used both Torch and TensorFlow.

C. Results

1) Misinformation detection: Table IV shows the obtained result on our experiments. The model with the best score is BERT with f1-score of 97.5% and ROC score of 99.6% which is higher than the state-of-the-art score. The second-best model is BiGRU with a difference of only 2.2% on f1-score and 15 times fewer parameters. This is therefore the model we will use later to pick out the tweets whose propagation we intend to predict in the network.

TABLE IV: Misinformation detection models scores. ST for sentence-transformer encoder model

Model	Parameters	F1-score	ROC
BiGRU	6,184,002	*0.953*	0.99
FastText	1,284,418	0.937	0.985
LogReg	**20,000**	0.931	0.985
NBSVM	**20,000**	*0.936*	0.986
ST+SVM	-	0.943	0.987
ST+XGB	-	0.939	0.983
BERT	109,136,642	**0.975**	**0.996**

2) Misinformation spreads: Figure 3 shows the results obtained in our experiment. Around 3 of the 10 most likely misleading tweets that a user will share are actually shared.

VI. Conclusion

A. Conclusions on the presented work

In this work, we propose a recurrent neural architecture with very few parameters to classify the information shared on social networks and to predict with some confidence which users will share this information.

The Covid-19 misinformation detection model was trained on the COVID- 19 Fake News Dataset. We propose to use the matrix factorization principle to offer tweet recommendations to users who have already agreed with the idea expressed in the tweet.

These two models blocks will be able to be used to propose alternatives of tweets to the users of the platform (the mistaken tweets detected on the network will be able to be distributed with those true, thus giving the possibility to the users better to be able to judge before sharing), to detect on people the virulent ones in the next minutes to limit the programming until mistaken information to which they adhere.

B. Perspectives

The proposed recommender model does not consider relationships between users, which could be a major asset since the information is transmitted from the creator of the tweet to his followers and then to the followers of his followers. Due to the limitation imposed by the Twitter API, we unfortunately could not access this information.

We could also increase the user interaction dataset. As we said above, we have opted for networks that do not have the same level of data.

Acknowledgment

This research was supported by the University of CY cergy Paris/ENSEA. We thank our Professor "Dimitris Kotzinos" who provided insight and judgement on our oral presentation that greatly assisted in the making of this paper.

References

[1] Heidi Oi-Yee Li et al. "YouTube as a source of information on COVID-19: a pandemic of misinformation?" In: *BMJ Global Health* 5 (2020).

[2] Yvonne Gangloff. *Youtube: video fact-checking comes to Europe*. Sept. 2020. URL: https://siecledigital.fr/2020/09/28/youtube-fact-check-europe/.

[3] Alice Huot. *How TikTok is fighting fake news about the Covid-19 vaccine*. Dec. 2020. URL: https://www.ladn.eu/media-mutants/reseaux-sociaux/vaccins-anti-covid-tiktok-lutte-fake-news/.

[4] Md. Rafiqul Islam et al. "Deep learning for misinformation detection on online social networks: a survey and new perspectives". In: *Social Network Analysis and Mining* 10 (2020).

[5] Liang Wu et al. "Misinformation in Social Media: Definition, Manipulation, and Detection". In: *SIGKDD Explor.* 21 (2019), pp. 80–90.

[6] Zhiwei Jin et al. "Detection and Analysis of 2016 US Presidential Election Related Rumors on Twitter". In: *SBP-BRiMS*. 2017.

[7] Kate Starbird et al. "Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing". In: *iConference 2014 Proceedings*. 2014.

[8] Suisheng Yu, Mingcai Li, and Fengming Liu. "Rumor Identification with Maximum Entropy in MicroNet". In: *Complex.* 2017 (2017), 1703870:1–1703870:8.

[9] Gisel Bastidas Guacho et al. "Semi-supervised Content-Based Detection of Misinformation via Tensor Embeddings". In: *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (2018), pp. 322–325.

[10] Adrien Friggeri et al. "Rumor Cascades". In: *ICWSM*. 2014.

[11] Sejeong Kwon and M. Cha. "Modeling Bursty Temporal Pattern of Rumors". In: *ICWSM*. 2014.

[12] Jooyeon Kim, Dongkwan Kim, and Alice H. Oh. "Homogeneity-Based Transmissive Process to Model True and False News in Social Networks". In: *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining* (2019).

[13] Alon Jacovi, Oren Sar Shalom, and Yoav Goldberg. "Understanding Convolutional Neural Networks for Text Classification". In: *BlackboxNLP@EMNLP*. 2018.

[14] Yi-Chin Chen, Zhao-Yang Liu, and Hung-Yu Kao. "IKM at SemEval-2017 Task 8: Convolutional Neural Networks for stance detection and rumor verification". In: **SEMEVAL*. 2017.

[15] Sarah A. Alkhodair et al. "Detecting breaking news rumors of emerging topics in social media". In: *Inf. Process. Manag.* 57 (2020), p. 102018.

[16] Jing Ma et al. "Detecting Rumors from Microblogs with Recurrent Neural Networks". In: *IJCAI*. 2016.

[17] Feng Qian et al. "Neural User Response Generator: Fake News Detection with Collective User Intelligence". In: *IJCAI*. 2018.

[18] Jing Ma, Wei Gao, and Kam-Fai Wong. "Detect Rumors on Twitter by Promoting Information Campaigns with Generative Adversarial Learning". In: *The World Wide Web Conference* (2019).

[19] Jure Leskovec, Lada A Adamic, and Bernardo A Huberman. "The dynamics of viral marketing". In: *ACM Transactions on the Web (TWEB)* 1.1 (2007), 5–es.

[20] Duncan J Watts and Peter Sheridan Dodds. "Influentials, networks, and public opinion formation". In: *Journal of consumer research* 34.4 (2007), pp. 441–458.

[21] Jacco Wallinga and Peter Teunis. "Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures". In: *American Journal of epidemiology* 160.6 (2004), pp. 509–516.

[22] P Alex Dow, Lada Adamic, and Adrien Friggeri. "The anatomy of large facebook cascades". In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 7. 1. 2013, pp. 145–154.

[23] Jure Leskovec, Lars Backstrom, and Jon Kleinberg. "Meme-tracking and the dynamics of the news cycle". In: *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2009, pp. 497–506.

[24] Cheng Li et al. "Deepcas: An end-to-end predictor of information cascades". In: *Proceedings of the 26th international conference on World Wide Web*. 2017, pp. 577–586.

[25] Qi Cao et al. "Deephawkes: Bridging the gap between prediction and understanding of information cascades". In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 2017, pp. 1149–1158.

[26] Mohammad Raihanul Islam et al. "Deepdiffuse: Predicting the'who'and'when'in cascades". In: *2018 IEEE International Conference on Data Mining (ICDM)*. IEEE. 2018, pp. 1055–1060.

[27] Jia Wang et al. "Topological recurrent neural network for diffusion prediction". In: *2017 IEEE International Conference on Data Mining (ICDM)*. IEEE. 2017, pp. 475–484.

[28] Zhitao Wang, Chengyao Chen, and Wenjie Li. "A sequential neural information diffusion model with structure attention". In: *Proceedings of the 27th ACM international conference on information and knowledge management*. 2018, pp. 1795–1798.

[29] Cheng Yang et al. "Multi-scale Information Diffusion Prediction with Reinforced Recurrent Networks". In: *IJCAI*. 2019.

[30] Parth Patwa et al. *Fighting an Infodemic: COVID-19 Fake News Dataset*. 2020. arXiv: 2011.03327 [cs.CL].

[31] Rabindra Lamsal. "Design and analysis of a large-scale COVID-19 tweets dataset". In: *Applied Intelligence* 51.5 (2021), pp. 2790–2804.

[32] Wenhui Wang et al. "MiniLM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers". In: *ArXiv* abs/2002.10957 (2020).

[33] Kyunghyun Cho et al. *On the Properties of Neural Machine Translation: Encoder-Decoder Approaches*. 2014. DOI: 10.48550/ARXIV.1409.1259. URL: https://arxiv.org/abs/1409.1259.

[34] Jacob Devlin et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2018. DOI: 10.48550/ARXIV.1810.04805. URL: https://arxiv.org/abs/1810.04805.

[35] Ruiqi Guo et al. "Accelerating Large-Scale Inference with Anisotropic Vector Quantization". In: *International Conference on Machine Learning*. 2020. URL: https://arxiv.org/abs/1908.10396.

[36] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[37] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: ACM, 2016, pp. 785–794. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785. URL: http://doi.acm.org/10.1145/2939672.2939785.

[38] Nils Reimers and Iryna Gurevych. "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Nov. 2019. URL: https://arxiv.org/abs/1908.10084.

[39] Arun S. Maiya. "ktrain: A Low-Code Library for Augmented Machine Learning". In: *arXiv preprint arXiv:2004.10703* (2020). arXiv: 2004.10703 [cs.LG].