

# BaitBuster: Destined to Save You Some Clicks

Md Main Uddin Rony <sup>§</sup>, Naeemul Hassan <sup>§</sup>, Mohammad Yousuf <sup>\*</sup>

<sup>§</sup>Department of Computer Science and Engineering, The University of Mississippi

<sup>\*</sup>Gaylord College of Journalism and Mass Communication, The University of Oklahoma

## ABSTRACT

The use of tempting headlines (clickbait) to allure readers has become a growing practice nowadays. The widespread use of clickbait risks the reader's trust in media. According to a study performed by Facebook, 80% users "preferred headlines that helped them decide if they wanted to read the full article before they had to click through". In this paper, we present a clickbait detection model which uses distributed subword embeddings and achieves an accuracy of 98.3%. Powered with the model, we build **BaitBuster**, a solution framework (social bot+browser extension), which not only detects clickbaits floating on the web but also provides brief explanation behind its action. Moreover, we study 1.67 million contents created by 153 media organizations and discover the relation between clickbait usage and media reliability.

## ACM Reference format:

Md Main Uddin Rony <sup>§</sup>, Naeemul Hassan <sup>§</sup>, Mohammad Yousuf <sup>\*</sup>. 2017. BaitBuster: Destined to Save You Some Clicks. In *Proceedings of Computation+Journalism Symposium, Illinois, USA, October 2017 (Computation+Journalism)*, 5 pages.  
<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

## 1 INTRODUCTION

*Facebook engineers panic, pull plug on AI after bots develop their own language*<sup>1</sup> – this headline and its similar versions, suggesting an apocalyptic situation, have misrepresented the actual facts but disrupted (shared more than 300K times <sup>2</sup> on Facebook alone) the social media. We call headlines like these as *Clickbait*. The term *clickbait* refers to a form of web content that employs writing formulas and linguistic techniques in headlines to trick readers into clicking links [5, 18], but does not deliver on promises <sup>3</sup>. Table 1 shows examples of some clickbait headlines. Media scholars and pundits consistently show clickbait content in a bad light, but the industry based on this type of content has been rapidly growing and reaching more and more people across the world [14, 21]. *Taboola*, one of the key providers of clickbait content, claims <sup>4</sup> to have doubled its monthly reach from 500 million unique users to 1 billion

<sup>1</sup><https://web.archive.org/web/20170731175117/http://bgr.com/2017/07/31/facebook-ai-shutdown-language/>

<sup>2</sup><http://bit.ly/2vgHnzz>

<sup>3</sup><https://www.wired.com/2015/12/psychology-of-clickbait/>

<sup>4</sup><https://www.taboola.com/ru/node/5046>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

*Computation+Journalism, October 2017, Illinois, USA*

© 2017 Copyright held by the owner/author(s).

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

in a single year from March 2015. The growth of clickbait industry appears to have clear impact on the media ecosystem, as many traditional media organizations have started to use clickbait techniques to attract readers and generate revenue. However, media analysts suggest that news media risk losing readers' trust and depleting brand value by using clickbait techniques that may boost advertising revenue only temporarily. According to a study performed by Facebook <sup>5</sup>, 80% users "preferred headlines that helped them decide if they wanted to read the full article before they had to click through". [15] shows that clickbait headlines lead to negative reactions among media users.

In this paper, we present a novel technique to detect clickbait headlines. We model the detection task as a supervised classification problem. Instead of following the traditional bag-of-words and hand-crafted feature set approaches, we adopt deep learning techniques that does not require feature engineering. Specifically, we use distributed subword embeddings [3, 13] to transform words into 300 dimensional embeddings. These embeddings are used to map sentences into vectors over which a softmax function is applied as a classifier. The learned classification model is used to build a clickbait solution framework, BaitBuster. It primarily has two components- a browser extension and a social bot. Unlike the existing clickbait detection browser extensions, it provides explanations of why a headline is a clickbait. Moreover, it also allows users to read the brief summary of the article without leaving the current page in case the user is interested. The social bot regularly publishes automatically generated report about contemporary clickbait articles. The objective of this bot is to fight against the rising number of malicious bots which breathe on clickbait, listicle <sup>6</sup> and fake contents. Details of BaitBuster is presented in section 6.

Table 1: Examples of clickbait headlines

No.	Headline
1	You Won't Believe What These Dogs Are Doing!
2	38 Things No Man Over 40 Should Own
3	Do this EVERY TIME you turn on your computer!
4	What OJ's Daughter Looks Like Now is Incredible!

To understand the relation between the practice of clickbait usage and media reliability, we study a large corpus of approximately 1.67 million contents created during the year of 2014–2016 by 68 mainstream media and 85 unreliable (vetted by multiple source) media organizations. We discover that an average unreliable media uses clickbaits about 15% more than an average print media. Details of the analysis is presented in later sections.

<sup>5</sup><https://tinyurl.com/ya2tykna>

<sup>6</sup><https://en.wikipedia.org/wiki/Listicle>

## 2 BACKGROUND

Even though *clickbait* is a relatively nascent term, its traces can be found in several journalistic concepts such as *tabloidization* and *content trivialization*. The linguistic techniques and presentation styles, employed typically in clickbait headlines and articles, derived from the tabloid press that baits readers with sensational language and appealing topics such as celebrity gossip, humor, fear and sex [18]. The Internet and especially the social media have made it easier for the clickbait practitioners to create, publish in a larger scale and reach a broader audience with a higher speed than before [12]. In the last several years, academicians and media studied this phenomenon from several perspectives.

**Clickbait– Properties, Practice and Effects:** There have been a small number of studies—some conducted by academic researchers and others by media firms—which examined correlations between headline attributes and degree of user engagement with content. Some media market analysts and commentators [7] discussed various aspects of this practice. However, no research has been found, which gauges the extents of clickbait practices by mainstream and alternative media outlets on the web.

Researchers at the University of Texas’s *Engaging News Project* [15] conducted an experiment on 2,057 U.S. adults to examine their reactions to clickbait (e.g., question-based headlines) and traditional news headlines in political articles. They found that clickbait headlines led to more negative reactions among users than non-clickbait headlines.

*Chartbeat*, an analytics firm that provides market intelligence to media organizations, tested 10,000 headlines from over 100 websites for their effectiveness in engaging users with content [4]. The study examined 12 ‘common tropes’ in headlines— a majority of them are considered clickbait techniques – and found that some of these tropes are more effective than others. Some media pundits interpreted the findings of this study as clickbaits being detrimental to traditional news brands.

*HubSpot* and *Outbrain*, two content marketing platforms that distribute clickbait contents across the web, examined millions of headlines to identify attributes that contribute to traffic growth, increased engagement, and conversion of readers into subscribers. The study suggested that clickbait techniques may increase temporary engagement [9], but an article must deliver on its promises made in headline for users to return and convert.

**Automated Clickbait Detection:** [1, 5, 19, 22] study automated detection of clickbait headlines using natural language processing and machine learning. [22] collects 10,000 headlines from *Buzzfeed*, *Clickhole*, and *The New York Times (NYT)* and uses Logistic Regression to create a supervised clickbait detection model. It assumes all *Buzzfeed* and *Clickhole* headlines as clickbait and all *NYT* headlines as non-clickbait. We would like to argue that it makes the model susceptible to personal bias as it overlooks the fact that many *Buzzfeed* contents are original, non-clickbaity and there is clickbait practice in *NYT* as well [10]. Moreover, *BuzzFeed*, and *NYT* usually write on very different topics. The model might have been trained merely as a topic classifier. [19] attempts to detect clickbaity Tweets in Twitter by using common words occurring in clickbaits, and by extracting some tweet specific features. [5] uses a dataset of 15,000 manually labeled headlines to train several supervised models for

**Table 2: Distribution of the Media Corpus**

Media	Category	Link	Video	Total
Mainstream (68)	Broadcast (43)	324028	32924	356952
	Print (25)	516713	14129	530842
Unreliable (85)	Clickbait (30)	371834	4099	375933
	Conspiracy (31)	309122	5841	314963
	Junk Science (6)	51923	649	52572
	Satire (18)	41046	151	41197
Total (153)		1614666	57793	1672459

clickbait detection. These methods heavily depend on a rich set of hand-crafted features which take good amount of time to engineer and sometimes are specific to the domain (for example, tweet related features are specific to Twitter data and inapplicable to other domains).

## 3 DATASET

We use two datasets in this paper. Below, we provide description of the datasets and explain the collection process.

**Headlines:** This dataset is curated by Chakraborty et al. [5]. It contains 32,000 headlines of news articles which appeared in ‘WikiNews’, ‘New York Times’, ‘The Guardian’, ‘The Hindu’, ‘BuzzFeed’, ‘Upworthy’, ‘ViralNova’, ‘Thatscoop’, ‘Scoopwhoop’, and ‘ViralStories’<sup>7</sup>. Each of these headlines is manually labeled either as a clickbait or a non-clickbait by at least three volunteers. There are 15,999 clickbait headlines and 16,001 non-clickbait headlines in this dataset. We used this labeled dataset to develop a clickbait classification model (details in Section 4). An earlier version of this dataset which had an even distribution of 7,500 clickbaits and 7,500 non-clickbaits was used in [1, 5].

**Media Corpus:** For large scale analysis, using Facebook Graph API<sup>8</sup>, we accumulated all the Facebook posts created by a set of mainstream and unreliable media within January 1<sup>st</sup>, 2014 and December 31<sup>st</sup>, 2016. The mainstream set consists of the 25 most circulated print media<sup>9</sup> and the 43 most-watched broadcast media<sup>10</sup> (according to Nielson rating [6]) in the United States. The unreliable set is a collection of 85 conspiracy, clickbait, satire and junk science based media organizations. The category of each unreliable media is cross-checked by two sources [11, 24]. Overall, we collected more than 2 million Facebook posts. A post may contain a photo or a video or a link to an external source. In this paper, we limit ourselves to the link and video type posts only. This reduces the corpus size to 1.67 million. For each post, we collect the headline (title of a video or headline of an article) and the status message. For a collection of 191,540 link type posts, we also collected the bodies of the corresponding news articles. Table 2 shows distribution of the corpus. Number of media organizations in each category is listed in parentheses.

<sup>7</sup><https://github.com/bhargaviparanjape/clickbait/tree/master/dataset>

<sup>8</sup><https://developers.facebook.com/docs/graph-api>

<sup>9</sup>[https://en.wikipedia.org/wiki/List\\_of\\_newspapers\\_in\\_the\\_United\\_States](https://en.wikipedia.org/wiki/List_of_newspapers_in_the_United_States)

<sup>10</sup>[www.indiewire.com/2016/12/cnn-fox-news-msnbc-nbc-ratings-2016-winners-losers-1201762864/](http://www.indiewire.com/2016/12/cnn-fox-news-msnbc-nbc-ratings-2016-winners-losers-1201762864/)

## 4 CLICKBAIT DETECTION

The key component of BaitBuster is the clickbait detection model. In this section, we formally define the detection problem, briefly describe the techniques used, and present evaluation of the model’s performance.

### 4.1 Problem Definition

We define the clickbait identification task as a supervised binary classification problem where the set of classes is  $C = \{\text{clickbait}, \text{non-clickbait}\}$ . Formally, given  $\mathcal{X}$ , a set of all sentences, and a training set  $\mathcal{S}$  of labeled sentences  $\langle s, c \rangle$ , where  $\langle s, c \rangle \in \mathcal{X} \times C$ , we want to learn a function  $\gamma$  such that  $\gamma : \mathcal{X} \rightarrow C$ , in other words, it maps sentences to  $\{\text{clickbait}, \text{non\_clickbait}\}$ .

A traditional approach in text classification is to use *bag-of-words* (BOW) model to transform text into feature vectors before applying learning algorithms. However, inspired by the recent success of deep learning methods in text classification, we use distributed subword embeddings as features instead of applying BOW model. Specifically, we use an extension, named as Skip-Gram<sub>sw</sub>, of the continuous *skip-gram* model [17], which leverages subwords (substring of a word) to compute word embeddings [3]. Given a word, *skip-gram* wants to maximize the correct prediction of its context. Skip-Gram<sub>sw</sub> works in a slightly different way. Rather than treating each word as a unit, it breaks down words into subwords and wants to correctly predict the context subwords of a given subword. This extension allows sharing the representations across words, thus allowing to learn reliable representations for rare words. Consider the following example.

*Example 4.1.* “the quick brown fox jumped over the lazy dog”-take the word “quick” as an example. Assuming subword length as three, the subwords are-  $\{qui, uic, ick\}$ . Skip-Gram<sub>sw</sub> model learns to predict *qui*, *ick* in the context given *uic* as the input. Assuming “sick” as a rare word (absent in the vocabulary), the embeddings of *ick* and *sic* are used to learn an approximate representation of it.

Using a neural network architecture, Skip-Gram<sub>sw</sub> learns the mapping between the output and the input. The weights to the hidden layer form the vector representations of the subwords. The embedding of a word is formed by the summation of the vector representations of its subwords. We take the average of the embeddings of words present in a sentence to form the representation of the sentence. These sentence representations are used to train a linear classifier. Further details of the model can be found in [3, 13].

### 4.2 Evaluation

Based on Skip-Gram<sub>sw</sub>, we propose five different models. The Headlines dataset is used to evaluate the classification models. A 10-fold cross-validation is performed. We repeat each experiment 5 times to avoid any random behavior. The best model achieves an accuracy of 98.3% which is significantly better than the accuracy (93%) reported in [5]. We also compare our models with other existing works in terms of precision, recall, and other measures. A detailed description of the five models, experiment setup, and performance evaluation can be found in [20].

**Table 3: Amount of clickbaits in various media**

Media	Category	Clickbait	Non-clickbait	Clickbait (%)
Mainstream	Print	128022	402820	24.12
	Broadcast (news)	28543	99314	22.32
	Broadcast (non-news)	141209	87886	61.64
Unreliable	Clickbait	172271	203662	45.82
	Conspiracy	90389	224574	28.7
	Junk Science	23637	28935	44.96
	Satire	21798	19399	52.91

## 5 CLICKBAIT AND MEDIA RELIABILITY

We use the Media Corpus dataset to study the relation between clickbait practice and media reliability. We apply the Skip-Gram<sub>sw</sub> model on the headlines to measure clickbaitiness of the posts. Table 3 shows amounts of clickbaits and non-clickbaits in the headlines of mainstream and unreliable media. By and large, unreliable media uses clickbaits more than the mainstream media. Out of 784,665 posts by unreliable media, 308,095 (39.26%) have clickbait headlines. In print media, the ratio is 24.12%, about 15% less. The Media Corpus has 43 broadcast media. We manually categorize them into news oriented broadcast media (e.g. *CNN*, *NBC*, etc.) and non-news (lifestyle, entertainment, sports, etc.) broadcast media (e.g. *HGTV*, *E!*, etc.). There are 6 news oriented broadcast media and 37 non-news broadcast media in the corpus. We find that the ratio of clickbait and non-clickbait is 61.64% in non-news type broadcast media whereas it is only 22.32% (close to print media) in news oriented media.

**Table 4: A list of the top-20 clickbait practitioners**

Name	Category	Clickbait	Non-clickbait	Clickbait(%)
VH1	Broadcast (non-news)	13760	1339	91.13
AmplifyingGlass	Unreliable	692	71	90.69
MTV	Broadcast (non-news)	42313	4492	90.4
ClickHole	Unreliable	8250	930	89.87
Reductress	Unreliable	3984	484	89.17
Bravo TV	Broadcast (non-news)	8263	1242	86.93
Food Network	Broadcast (non-news)	2990	492	85.87
OWN	Broadcast (non-news)	474	118	80.07
E!	Broadcast (non-news)	24501	6167	79.89
Food Babe	Unreliable	2387	638	78.91
Chicks on the Right	Unreliable	14185	4977	74.03
HGTV	Broadcast (non-news)	1886	681	73.47
TLC	Broadcast (non-news)	787	296	72.67
RealFarmacy	Unreliable	2237	884	71.68
BET	Broadcast (non-news)	12894	5390	70.52
Nick at Nite	Unreliable	179	88	67.04
Tea Party News (TPNN)	Unreliable	4710	2356	66.66
WE TV	Broadcast (non-news)	943	478	66.36
Newslo	Unreliable	437	229	65.62
TV Land	Broadcast (non-news)	160	87	64.78

Table 4 shows the top-20 clickbait practitioners among the 153 media organizations of our Media Corpus dataset. For each organization, it shows the number of contents containing clickbait headlines, non-clickbait headlines, and the percentage of clickbaits. There are 9 unreliable media and 11 non-news broadcast media making into this list. No mainstream print or news oriented broadcast media made into the list.

## 6 BAITBUSTER

Standing on the clickbait classification model, we build BaitBuster, a clickbait solution framework, to improve the web surfing experience of general users. At this stage, the framework consists of a browser extension that identifies clickbaits present in a Facebook timeline, and a Facebook page which is administered by a bot through the Facebook API. The reason behind choosing Facebook is, according to a study performed by Pew Research, 44% of American adults get their news from Facebook [8]. Even though the framework is currently not available for public use, we plan to publish it before the CJ2017 symposium starts. Below, we present the architecture, design, and use case of BaitBuster.

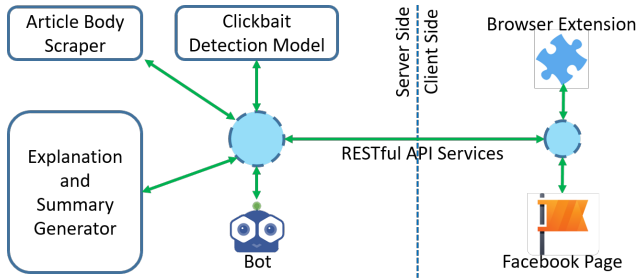


Figure 1: System architecture of BaitBuster

### 6.1 System Overview

Figure 1 shows the architecture of the BaitBuster framework. The browser extension monitors a user's Facebook news feed and alerts her if a post (link) contains a clickbait headline. In addition, it provides a brief explanation behind the decision which includes- what language features present in a headline makes it a clickbait, whether the headline represents the corresponding body fairly, and if the content is published from a controversial source. The extension also shows a brief summary of the corresponding article so that the user can get the gist without leaving the current page. Users can also provide feedback on the correctness of the classification done by BaitBuster. The social bot monitors which clickbait headlines are trending and publishes a brief report about them in a Facebook page through API. The goal of the bot is to discourage users from sharing clickbait contents by making them aware of their clickbaitiness.

### 6.2 Implementation

We implement the BaitBuster framework following the client-server system architecture model. The client and the server communicate with each other through a set of RESTful API services.

**6.2.1 Client.** This component has a JavaScript based browser extension and a bot powered BaitBuster Facebook page. The extension scans the Document Object Model (DOM) of the current page, identifies the anchor elements (`<a href=...>`), and sends the anchor texts and the corresponding URLs to the server side using POST request. The server processes the request (details in section 6.2.2) and sends response back to the client. Then the extension creates a new DOM object with the clickbait decision for



Figure 2: BaitBuster identifies clickbait post in Facebook



Figure 3: Explanation of the decision and content Summary

each anchor and inserts the object above the corresponding anchor element. Figure 2 and 3 show the graphical user interface (GUI) of the extension. The example post in the figure contains a clickbait headline. For the posts which are not classified as clickbait, BaitBuster just shows options for providing feedback, so that users can inform about false negatives. The BaitBuster Facebook page is administered by a bot using Facebook API (details in section 6.2.2).

**6.2.2 Server.** The server side is developed using Python and the Flask micro-framework. It has several components. The *Clickbait Detection Model* is explained in section 4. We briefly describe the other components below.

**Article Body Scraper:** BaitBuster uses a python package called *Newspaper*<sup>11</sup> to extract the headline and the body of an article given the URL to it. This information is used to measure whether a headline fairly represent its body or not, and to generate a brief summary of the article. The extracted data is stored in a database.

**Explanation and Summary Generator:** We prepare a list of 1000 most frequent n-grams ( $n = 3$ ) present in the clickbait examples of the Headlines dataset. This component detects if any of the n-grams is present in the requested (by the client side browser extension) post's headline. It also identifies if the post is created by a controversial (fake, conspiracy, unreliable, rumor, etc.) source. For the controversiality information, it leverages a professionally curated list of 834 online sources<sup>12</sup>. Moreover, this component measures how fairly a headline represents its body. We use Gensim's<sup>13</sup> TextRank [2, 16] based summarizer to extract the summary of a body. Then, cosine similarity is used to measure the similarity between the summary and the corresponding headline. Rather than simply showing the cosine similarity score, this component provides a comparison with respect to an average non-clickbait article. We measure the average of the headline-body similarities of the non-clickbait examples of the Headlines dataset. If a requested post's headline-body similarity is lesser than the average headline-body similarity, this component reports that to the user as depicted in Figure 3. As the source of this post is not present in the curated list, it's not showed on the GUI.

**Bot:** Activities of the browser extension are logged in a database. These data allows us to know, for any time interval, the most viewed

<sup>11</sup><https://github.com/codelucas/newspaper>

<sup>12</sup><http://www.opensources.co/>

<sup>13</sup><https://radimrehurek.com/gensim/>

clickbait posts and their source controversiality. A program automatically generates a small report (by filling up a template) with the most viewed clickbait posts on a day including their source identity and controversiality. Using Facebook API, it publishes the report to the client side page on a daily basis. As numerous malicious bots are spreading disinformation across the web [23], we believe this is a small step towards fighting disinformation using benevolent bots. Even though the bot currently forewarns about clickbait posts only, our goal is to cover fake news and misleading contents in future.

### 6.3 Use Case of the Browser Extension

A user installs the BaitBuster browser extension from the marketplace. When she visits Facebook, she sees a red button with title "Potential Clickbait! - Click here to read more" on top of any clickbait links (Figure 2). If she clicks the button, a modal containing supplementary explanation and the summary of the corresponding body opens up (Figure 3). If she thinks the clickbait classification is correct, she can click the thumbs-up button. She can use the thumbs-down button if the extension misclassified a link. Collected user feedbacks are used to retrain and improve the performance of the classification model.

### 6.4 Comparison with Existing Solutions

There have been several attempts to limit clickbaits using browser extensions. For instance, *B.S. Detector*<sup>14</sup>, *Clickbait Killer*<sup>15</sup>, *Check This by MetaCert*<sup>16</sup> maintain an aggregated list of sources and check web contents against the set of sources. One limitation of this approach is it doesn't allow checking contents which hasn't already been checked by the aggregated list. *Stop Clickbait* [5] and *This is Clickbait*<sup>17</sup> use supervised models to check clickbaits. However, they don't provide any explanation behind the model's decision. We argue that only checking the headline of an article is not sufficient to determine whether its clickbait or not; the body needs to be checked and collocated with the headline as well. According to our knowledge, only BaitBuster provides deep learning powered classification and supplements it with explanation and summary by leveraging the headline-body relation.

## 7 CONCLUSION AND FUTURE WORK

In this paper, we present a distributed sub-word embedding based clickbait detection technique. The classification model outperforms existing methods in terms of accuracy. We apply the model on a large corpus of media contents to uncover relation between clickbait practice and media reliability. A solution framework, BaitBuster, is being developed to improve the web using experience with the help of a browser extension and a bot. In future, we want to investigate more thoroughly how the body of an article can be leveraged to improve the clickbait classification model. We also plan to investigate the relation between fake news and the clickbait practice in future.

## REFERENCES

- [1] Ankesh Anand, Tanmoy Chakraborty, and Noseong Park. 2016. We used Neural Networks to Detect Clickbaits: You won't believe what happened Next! *arXiv preprint arXiv:1612.01340* (2016).
- [2] Federico Barrios, Federico López, Luis Argerich, and Rosa Wachenchauzer. 2016. Variations of the similarity function of textrank for automated summarization. *arXiv preprint arXiv:1602.03606* (2016).
- [3] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2016. Enriching Word Vectors with Subword Information. *arXiv preprint arXiv:1607.04606* (2016).
- [4] Christopher Breau. 2015. You'll Never Guess How Chartbeat's Data Scientists Came Up With the Single Greatest Headline. <http://blog.chartbeat.com/2015/11/20/youll-never-guess-how-chartbeats-data-scientists-came-up-with-the-single-greatest-headline/>. (2015).
- [5] Abhijnan Chakraborty, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. 2016. Stop Clickbait: Detecting and preventing clickbaits in online news media. In *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on*. IEEE, 9–16.
- [6] John Eggerton. 2016. FCC: Nielsen DMAs Still Best Definition of TV Market. Broadcasting & Cable. (2016).
- [7] Frederic Filloux. 2016. Clickbait is devouring journalism but there are ways out. <https://qz.com/648845/clickbait-is-devouring-journalism-but-there-are-ways-out/>. (2016).
- [8] Jeffrey Gottfried and Elisa Shearer. 2016. News Use Across Social Media Platforms 2016. <http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/>. (2016).
- [9] Hubspot and Outbrain. [n. d.]. DATA DRIVEN STRATEGIES FOR Writing Effective Titles & Headlines. <http://cdn2.hubspot.net/hub/53/file-2505556912-pdf/DataDrivenStrategiesForWritingEffectiveTitlesAndHeadlines.pdf>. ([n. d.]).
- [10] Nathan Hurst. 2016. *To clickbait or not to clickbait? An examination of clickbait headline effects on source credibility*. Ph.D. Dissertation. University of Missouri-Columbia.
- [11] informationisbeautiful.net. 2016. Unreliable/Fake News Sites & Sources. <https://docs.google.com/spreadsheets/d/1xDDmbr54qzG8wUrRdxQlc1dixSIYqQUaXVZBqsJs>. (2016).
- [12] Mathew Ingram. 2014. The internet didn't invent viral content or clickbait journalism – there's just more of it now, and it happens faster. <http://bit.ly/1IPnfVv>. (2014).
- [13] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016. Bag of Tricks for Efficient Text Classification. *arXiv preprint arXiv:1607.01759* (2016).
- [14] MARK S. LUCKIE. 2015. ADELE AND THE DEATH OF CLICKBAIT. <http://www.niemanlab.org/2015/12/adele-and-the-death-of-clickbait/>. (2015).
- [15] Joshua M. Scacco and Ashley Muddiman. 2016. INVESTIGATING THE INFLUENCE OF "CLICKBAIT" NEWS HEADLINES. <https://engagingnewsproject.org/wp-content/uploads/2016/08/ENP-Investigating-the-Influence-of-Clickbait-News-Headlines.pdf>. (2016).
- [16] Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing Order into Text.. In *EMNLP*, Vol. 4. 404–411.
- [17] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.
- [18] Dolors Palau-Sampio. 2016. Reference press metamorphosis in the digital context: clickbait and tabloid strategies in Elpais. com. *Communication & Society* 29, 2 (2016).
- [19] Martin Potthast, Sebastian Köpsel, Benno Stein, and Matthias Hagen. 2016. Clickbait detection. In *European Conference on Information Retrieval*. Springer, 810–817.
- [20] Md Main Uddin Rony, Naeemul Hassan, and Mohammad Yousuf. 2017. Diving Deep into Clickbaits: Who Use Them to What Extents in Which Topics with What Effects?. In *2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM*.
- [21] Chris Sutcliffe. 2016. Can publishers step away from the brink of peak content? <https://www.themediabriefing.com/article/can-publishers-step-away-from-the-brink-of-peak-content/>. (2016).
- [22] Abhishek Thakur. 2016. Identifying Clickbaits Using Machine Learning. <https://www.linkedin.com/pulse/identifying-clickbaits-using-machine-learning-abhishek-thakur>. (2016).
- [23] Samuel C Woolley and Philip N Howard. [n. d.]. *Computational Propaganda Worldwide: Executive Summary*. Technical Report.
- [24] Melissa Zimdars. 2016. My 'fake news list' went viral. But made-up stories are only part of the problem. <https://www.washingtonpost.com/posteverything/wp/2016/11/18/my-fake-news-list-went-viral-but-made-up-stories-are-only-part-of-the-problem/>. (2016).

<sup>14</sup><http://bsdetecter.tech/>

<sup>15</sup><https://chrome.google.com/webstore/detail/clickbait-killer/bophlajogieadgcmhpbhmfneoninfj>

<sup>16</sup><https://chrome.google.com/webstore/detail/check-this-by-metacert/felmjclcdjapolhjmblbmfekjaojfbn>

<sup>17</sup><https://chrome.google.com/webstore/detail/this-is-clickbait/ppklhldfnadnlnlnenceabhlhdpnfajm?hl=en>