Examining Clickbait Through a Fake News Predictor

Team 10, The Data Open, Fall 2021

Alan Wu, Milo Knell, Forrest Bicker, David Chen

1 Executive Summary

Upworthy.com is a large media company which takes existing content and shares it on their website by changing the title and image. They optimize “packages,” a title and an image, by performing A/B tests to find what makes people click and share most. Even though they produce no original content, they have 12 times as many Facebook likes on average than other news companies (Fitts).

They have been accused of producing clickbait articles, with titles and images designed to maximize clicks and interactions - possibly at the cost of journalistic integrity. Due to bad press and plummeting page views, they took drastic action and rethought their business model. In February 2015 they hired a new chief story officer, Amy O'Leary, who pivoted the focus to original reporting and less clickbait (Sanders 2017). Later, in March 2015, they apologized for the influence of their clickbait and pledged to change (O'Reilly 2015).

Given Upworthy’s interesting history with clickbait, in our research, we propose a method to measure whether an article is clickbait. We theorize that fake news tends to have features similar to clickbait because it aims to pull in viewers. We trained a classifier for fake news on external data that only examines lexical features. Applying the classifier on Upworthy’s dataset of headlines, we found that fake news predicted clickbait more accurately than click rate alone. We did this by examining the percentage of daily articles classified as fake news, and found that that number dropped at nearly the same time as Upworthy’s pledge to change its editorial practices under the new management of Amy O’Leary. We found that predicted fake news is a good proxy to examine clickbait that avoids the influence of confounding variables like overall business performance and external factors that are not accounted for in the Upworthy data.

We next examined Upworthy’s business practices, their stance on clickbait and how this changed through time. We measured what we call an “override,” where Upworthy chooses to display a headline on their website even though it performed worse in tests. We found that before February 2015, when we observed the drop in fake news due to changed editorial practices, Upworthy tends to override for titles that appear more like clickbait. After this period they were nearly equal.

This analysis of clickbait, using fake news as a proxy, is significant because we found that clickbait tends to use more extreme emotional language (very positive or negative). This is potentially harmful because studies have shown that seeing extreme emotional language leads to worse mental health and emotional wellbeing across society.

2 Topic Question

We seek to answer a series of three questions:

1. Does fake news correlate with clickbait?
2. What is Upwrothy's stance on producing clickbait and how has it changed over time?
3. What is the impact of clickbait on the consumers of media?

3 Technical Exposition

3.1 Data

In addition to the Upworthy's A/B test dataset, we used a fake news dataset for training a fake news classifier. We used hand-made features instead of word-embeddings to prevent overfit.

3.1.1 Upworthy Dataset

Before we began our analysis of the Upworthy's dataset of 150,817 packages across 32,487 A/B tests, we first decided to filter our dataset to satisfy the following data integrity requirements: 1) Since we were specifically interested in examining the effect headlines have, we began by dropping all tests which did not control for the image. We did this because variation in images between A/B tests would be a confounding variable to our investigation into headlines, resulting in behavior in our dependent variables that would not be explainable exclusively by our independent variable of headline. 2) To ensure a sufficiently large sample size was used in the A/B tests we were analyzing for more robust analysis, we also removed all tests where any packages contained fewer than 100 impressions. This left us with an image-controlled dataset of 96,018 packages across 20,965 A/B remaining tests.

3.1.2 Fake News Dataset

The fake news dataset is combined from two sub-datasets and found in Horne 2017. The first dataset consists of 101 articles (53 real, 48 fake) that were gathered in 2016 before the election. Articles were selected by high engagement and split into real and fake. They then did filtering to ensure the quality of the data. The second dataset was gathered by the authors of the paper, and it consists of 150 samples (75 real and 75 fake) scraped from a known list of real and fake news sources. Opinion pieces were filtered out of both datasets. The result is one dataset of 251 headlines: 128 real and 123 fake.

3.1.3 Feature Extraction

To translate the qualitative text of both the Upworthy and fake news headlines into quantitative metrics, we featurized the text with a combination of several natural language processing techniques. We chose to manually engineer features as opposed to using word embeddings, because word embeddings often lead to overfitting, whereas feature engineering is more resilient to overfitting because it is deterministic and independent of sample distribution. We use feature engineering to prevent overfitting, which is especially important given the small size of the fake news dataset we used. We used feature engineering primarily drawn from Horne 2019. Our featurizing has 123 features, the full list of which is available in Appendix A. These features generally include complexity with features like word count, average word length, and various complexity measures SMOG readability index or the Flesch–Kincaid readability test. We also used psychology features like emotion analysis from Linguistic Inquiry and Word Count dictionaries (LIWC) and sentiment analysis from Valence Aware Dictionary and sentiment Reasoner (VADER). The final set of features is stylistic, with features like part of speech or punctuation counts (e.g. exclamation marks). The end result is a 123-dimensional vector quantitatively featurizing the headline texts.

3.2 Fake News Prediction

In order to measure clickbait, we set about to make a classifier that is superior to simply regressing the click rate of an article. We decided to use fake news derived from only lexical features since fake news often has flashy titles designed to pull in users, which behave very similar to clickbait.

3.2.1 Fake News Classifier

Using the featurization proposed in the previous section, we trained a linear SVM using the liblinear implementation with a squared hinge loss and l2 penalty. We used a tolerance of 1e-4, 1000 max iterations, and a 5 cross fold validation 10 step grid search over C values between 1e-4 and 1e+4, maximizing accuracy. The best C value was C=0.005995 with a test accuracy of 80.1% averaged across the 5 folds.

While other works have achieved accuracies as high as 98% on similar tasks with more data (such as LIAR or ISOT) (Hakak et al 2021), we believe that such high accuracies are the result of strong overfitting to the dataset. Humans cannot predict fake news with 100% accuracy on lexical features alone (i.e without prior knowledge if a claim is true or not). By opting to use a small but well maintained dataset we achieve a lower accuracy, but our model would be more realistic and less prone to finding external features that improve accuracy on in-domain data but fail to generalize out-of-domain.

3.2.2 Validation of Fake News Model

To validate this model we must show that it models clickbait better than click rate by removing confounding variables to present a better metric.

We first want to use a statistical test to show that click rate and predicted fake news label are strongly correlated. In this case our null hypothesis is that there is no correlation between click rate and predicted fake news.

Since the fake news label is ordinal data (it has categories with an internal order), the Spearman rank correlation is more appropriate than Pearson's R. To apply Spearman's test the data must be continuous or ordinal. That is, it either has continuous values (like click rate) or ordinal values (like fake news prediction). It shouldn't only contain categorical data but not ordinal data. Thus Spearman is an appropriate test in this instance.

Performing Spearman rank on fake news predicted label vs click rate yields a correlation of $r = 0.131$and a p value of $p = 0.0$(a p value less than 2.225e-308, which is the minimum value Python can store so it gets rounded to 0). This is a significant result at $\alpha = 0.05$ because $p << \alpha$ and we also have a moderate effect size, so we reject the null hypothesis that there is no relationship between predicted label and click rate and can therefore conclude that there is a relationship.

Second, we must show that even though they are strongly correlated, predicted fake news labels are a superior metric than click rate for measuring clickbait. We can do this by examining daily averaged click rate and the daily percent predicted fake news, as shown in Figure 1.

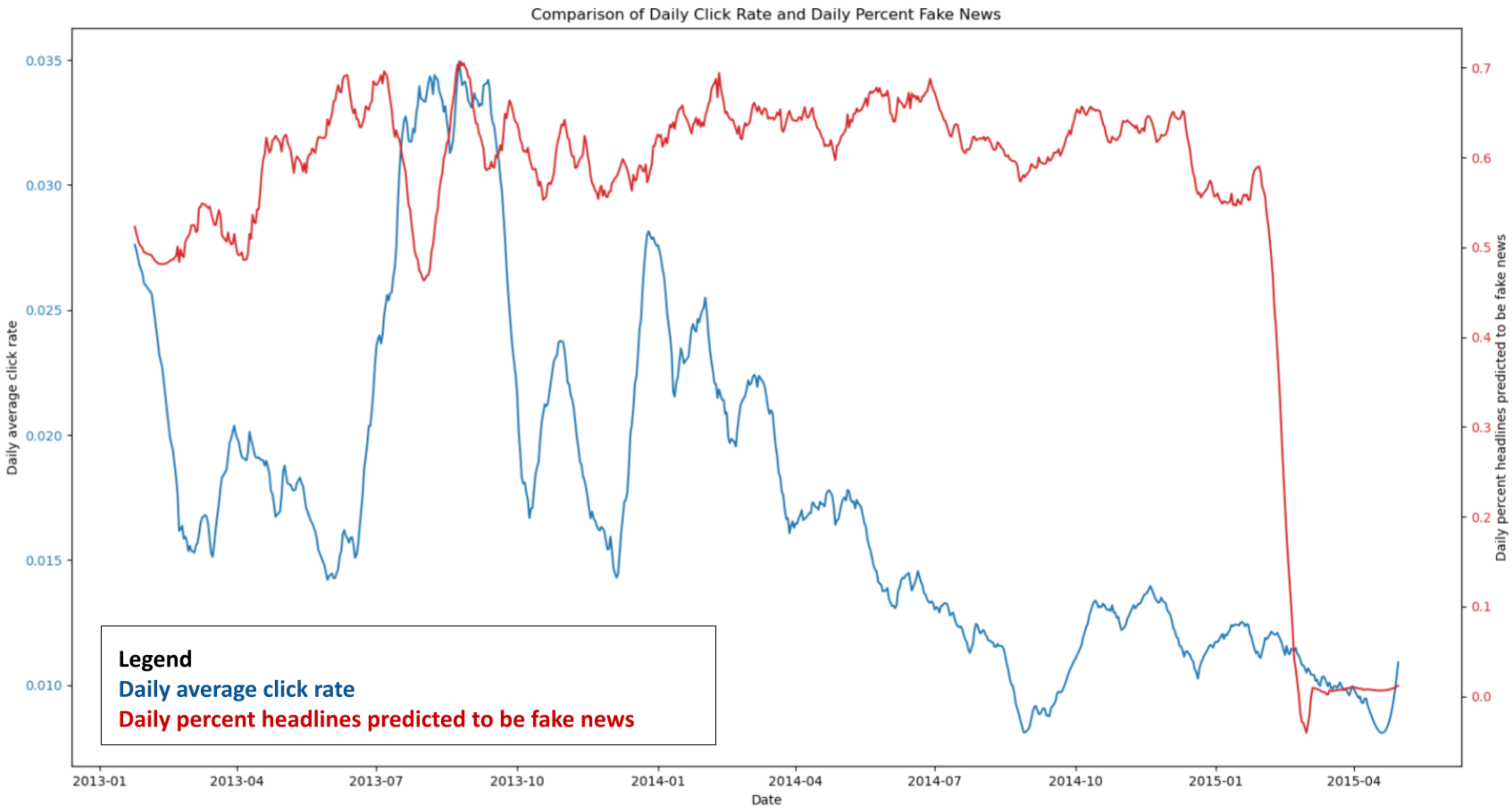

Figure 1: Comparison of Daily Click Rate (blue) and Daily Percent Predicted Fake News (red)

These lines are both smoothed using the Savitzky-Golay filter which interpolates using a low degree polynomial to smooth data without distorting tendencies. We used a window size of 41 and a polynomial of degree 3.

We can clearly see the strong drop in predicted fake news between February 11th and 15th in 2015. In February 2015, a new chief story officer was hired, Amy O'Leary, who hired more writers and changed the focus to show less clickbait news and do more original reporting (Sanders 2017). As we can see, **the percentage of predicted fake news drops nearly immediately as the journalistic practices change.** This is a strong validation for the model, as it models this drop while click rate does not.

An additional analysis we can perform is to show how date is a confounding variable on click rate but not on the predicted fake news label. This can be done by comparing date (quantified as integer counting days since the first day in the dataset, January 24th 2013) against click rate and the predicted fake news label. We perform this analysis only before the drop (February 11th 2015) to avoid the effect on only predicted fake news. We use Spearman rank where predicted fake news is ordinal while date and click rate are continuous. We have the following r and p values:

	Spearman's r (correlation)	p-value
Date vs Click Rate	-0.552	1.01e-58
Date vs Predicted Label	0.108	3.80e-3

This table shows that both click rate and predicted fake news label are significant at a $\alpha = 0.05$ margin, and that click rate is much more strongly correlated with date than predicted label, using the r value as the effect size. Click rate is 5.11 times more correlated than the predicted label. This shows how date is a confounding variable in click rate that is less impactful in fake news prediction.

With these results, we can argue that fake news **causes** higher click rate because our data satisfies the following conditions:

1. Correlation: Our Spearman rank analysis rejected the null hypothesis and proved statistical correlation between click rate and fake news
2. Temporality: We know that formulation of the headline (and therefore fake news labeling) necessarily always precedes experimental click
3. Experimental design: We are using Upwrothy's dataset of A/B tests, where treatments are randomly applied to randomly selected members of a population. This procedure satisfies the criteria for experimental design.
4. Non-supurious: We believe that the relationship between fake news and click rate is not caused by any other variables because we control for variation in all independent variables aside from headline selection by filtering our dataset to only include tests where headline is the same.

3.2.3 Summary

In this section we have shown that fake news prediction is a better proxy for clickbait than click rate. That is, click rate does not as accurately model clickbait as the fake news predictions we have proposed.

3.3 Business Behavior Analysis

Using the fake news predictor as a powerful proxy to represent clickbaits, we can examine Upworthy's business behavior in two characteristic periods. From Section 3.2 we see a characteristic pattern in Upworthy's clickbait production: the dramatic drop around February 15th, 2015. The clickbait production is estimated by fake news prediction, whose validity is shown in Section 3.2.

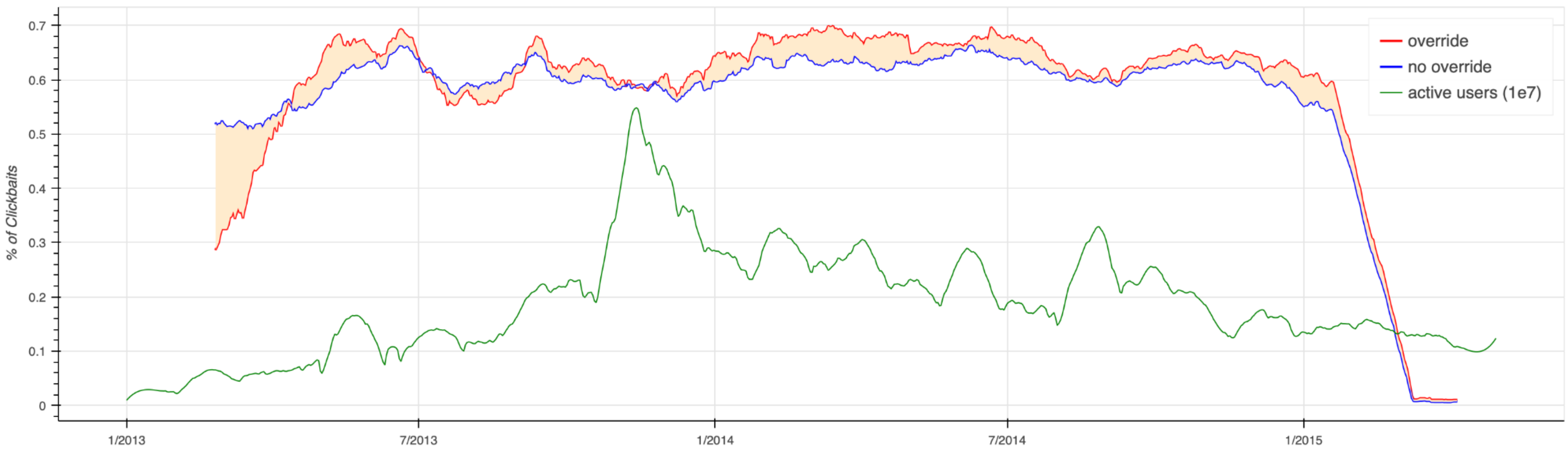

Figure 2: Red & Blue: The percentage of clickbaits produced each day, grouped by override. Green: Active users of Upworthy (1e7)

To examine business behavior, we focus on two aspects: 1) the percentage of clickbaits produced per day and 2) the overriding behavior (winner not equals first place) of Upworthy.

3.3.1 Behavior Before Feb 15th, 2015

Before the dramatic drop on Feb 15th, the percentage of clickbaits produced daily remained at a high level (~0.62). The active users kept growing and spiked on Nov 13. As the user count started to decrease, we spotted that the percentage of clickbaits in editor-overridden tests consistently surpassed the not overridden tests (Figure 2), which indicates that Upworthy was trying to select more clickbait-ish packages than what their A/B tests suggested. This behavior might be driven by the decreasing business performance, as similar patterns appear right after the drops in user counts, like in June 2013 and September 2014.

3.3.2 Behavior After Feb 15th, 2015

In February 2015, a new chief story officer, Amy O'Leary, came to Upworthy. She hired more writers and pledged to change the focus to showing less clickbait news and more original and trustworthy reporting (Sanders 2017). This historical event is reflected in the dramatic drop in clickbait percentage in Figure 2. After that, the percentage of clickbaits produced daily stayed low, which indicates that Upworthy generally stopped producing clickbait-ish titles in their packages.

3.3.3 Summary

Empowered by our fake news detector as an effective proxy to examine clickbaits, we drew insights from the clickbait production of Upwothy's packages. The behavior analysis confirms that Upworthy had produced an abundance of clickbaits that disrupted the internet, but it lived up to its pledge to correct it's business focus. For future works, this method of analysis could be applied to other news companies like Buzzfeed and Thought Catalog, to draw insights on their business behaviors over time.

3.4 Sentiment Factor Analysis On Clickbait

3.4.1 Analysis of VADER Sentiment

We set out to analyze the sentiment present in clickbait headlines. A component of the featurization is sentiment using the popular VADER library. We decided to measure the extreme emotions by defining *extreme = vader positive + vader negative*. This gives a measure of the emotional extremes present in a headline.

Our intuition was that clickbait would tend to contain more extreme vader emotions, so we applied Spearman Correlation to the extreme vader sentiment feature and the click rate. Fake news prediction is ordinal and click rate and vader extreme are both continuous.

	Spearman's r (correlation)	p-value
Vader Extreme vs Click Rate	0.0244	4.46e-14
Vader Extreme vs Fake News Prediction	0.0969	3.74e-199

Both click rate and fake news (which we have shown in Section 3.2, to be a superior proxy for clickbait) have a significant p-value and a positive correlation with extreme emotions. However, we note that the correlation between vader extreme and predicted fake news is 4 times higher than the correlation between vader extreme and click rate. Given that fake news is a superior proxy to examine clickbaits, this shows that clickbait contains more extreme emotions even when it is not as impactful in the click rate. Below we provide a few examples of the predicted VADER sentiments for articles.

Examples of VADER sentiment analysis

headline	Negativity	Neutrality	Positivity

The Ridiculous, Horrifying, Racist Reason Why Pot Is Illegal	0.734	0.266	0
The Completely Impractical, Woefully Contradictory Guide To Avoiding Rape	0.718	0.282	0
Did You Catch All These Events And News Stories In 2014?	0	1	0
The Happiest Man Alive Shares His Surprisingly Simple Secret	0	0.309	0.691
Stories Like His Restore My Faith In The Classic American Dream. How Awesome.	0	0.37	0.63

3.4.2 Social Impacts of Extreme Emotional Language

Our analysis shows that Upworthy is producing clickbait, which tends to have extreme emotional language. The impact of this is twofold: 1) First, studies have shown that excessive exposure to emotionally extreme media can increase anxiety, sadness, and depression (Piotrowski 2002). 2) Second, research has also demonstrated that extreme media have while also exacerbating personal worries and fears, even when those worries are unrelated to the content of the media (Johnston 2011).

Overall, seeing more extreme emotional language leads to worse mental health and emotional wellbeing across society. This impact might have applied to Upworthy viewers during the 2013-2015 timespan, and it might also extrapolate to consumers of any media outlet with similar clickbait-promoting tactics.

These impacts are further amplified by the current coronavirus pandemic: a study on Covid-19 headlines found that “around 52% of the news headlines evoked negative sentiments and only 30% evoked positive sentiments… Fear, trust, anticipation, sadness, and anger were the main emotions evoked by the news headlines” (Aslam 2020). During the period of the pandemic, much of the news headlines people see have extreme sentiment: this combines with the already large mental health and wellbeing consequences of the pandemic to increase anxiety and stress across society.

4 Conclusion

In summary, we built a classifier that accurately identifies headlines with lexicographical features typical of fake news. Our findings show that headlines that look like fake news are strongly correlated with clickbait, as evidenced by the moderate correlation with click rate and the presence of the drop in clickbait (with fake news as a proxy) detected when Upworthy’s editorial practices changed in February 2015. This better proxy for clickbaits allows us to analysis Upwothy’s business behavior, which confirms that Upworthy had produced an abundance of clickbaits, but it lived up to its pledge to correct it’s business focus. We additionally found that these clickbait titles are strongly correlated with extreme emotional language, hurting mental health and emotional wellbeing across society.

References

Hakak, Saqib et al. “An ensemble machine learning approach through effective feature extraction to classify fake news.” *Future Generation Computer Systems*, Volume 117, 2021, Pages 47-58, ISSN 0167-739X, https://doi.org/10.1016/j.future.2020.11.022.

O’Reilly, Laura. “Upworthy cofounder Peter Koechley apologizes for the clickbait 'monster' his site unleashed on the internet.” *Business Insider*, 2015, https://www.businessinsider.com/upworthy-co-founder-peter-koechley-at-guardian-changing-media-summit-2015-3

Fitts, Alex Sobel. “The king of content.” *Columbia Journalism Review*, https://archives.cjr.org/feature/the_king_of_content.php

Horne, Benjamin and Adali Sibel. “This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News.” *Rensselaer Polytechnic Institute*, 2017, https://arxiv.org/pdf/1703.09398.pdf .

Horne, Benjamin et al. “Robust Fake News Detection Over Time and Attack.” *ACM Transactions on Intelligent Systems and Technology*, Volume 11, Issue 1, 2019, Article No.: 7, pp 1–23, https://doi.org/10.1145/3363818.

Piotrowski CS & Brannen SJ (2002) Exposure, threat appraisal, and lost confidence as predictors of PTSD symptoms following September 11, 2001. *American Journal of Orthopsychiatry*, 72, 476-485.
https://pubmed.ncbi.nlm.nih.gov/15792033/

Johnston, W.M. and Davey, G.C.L. (1997). “The psychological impact of negative TV news bulletins: The catastrophizing of personal worries.” *British Journal of Psychology*, 88: 85-91. https://doi.org/10.1111/j.2044-8295.1997.tb02622.x

Appendix A: Full Feature List

These features are from Horne 2019, however we removed event detection because we found it to be unstable and cause errors. The original GitHub can be found at https://github.com/BenjaminDHorne/NELAFeatures, and our modified version is included in the code.

quotes	$	flesch_kincaid_grade_level
exclaim	'	smog_index
allpunc	(	coleman_liau_index
allcaps	)	lix
stops	,	bias_words
CC	--	assertatives
CD	.	factives
DT	:	hedges
EX	``	implicatives
FW	funct	report_verbs
IN	pronoun	positive_opinion_words
JJ	ppron	negative_opinion_words
JJR	i	tentat
JJS	we	certain
LS	you	vadneg
MD	shehe	vadneu
NN	they	vadpos
NNS	ipron	wneg
NNP	article	wpos
NNPS	verb	wneu
PDT	auxverb	sneg
POS	past	spos
PRP	future	sneu
PRP$	adverb	swear

RB	preps	affect
RBR	conj	posemo
RBS	negate	negemo
RP	quant	anx
SYM	number	anger
TO	cogmech	sad
UH	insight	HarmVirtue
WP$	cause	HarmVice
WRB	discrep	FairnessVirtue
VB	incl	FairnessVice
VBD	excl	IngroupVirtue
VBG	assent	IngroupVice
VBN	nonfl	AuthorityVirtue
VBP	filler	AuthorityVice
VBZ	ttr	PurityVirtue
WDT	avg_wordlen	PurityVice
WP	word_count	MoralityGeneral