Trends in Business Analytics

Cataneo & Co

Classifying Fake News

Business Question

The spread of misinformation threatens businesses and their reputations worldwide. A 2019 report from the University of Baltimore estimates that companies spend over $9.5B to combat the negative impact of misinformation on their brands annually[1]. To limit the spread of misinformation and mitigate its negative impact, companies need to find and debunk fake news as quickly as possible. For that reason, we explore the ability of data and analytics to address this issue. Specifically, we aim to answer the following questions:

- Can we build a classification model using data collected from the internet to predict whether or not a publication is fake news?
- How well does the model classify news as real or fake?
- How could companies implement a predictive model to mitigate the negative impact of misinformation on their business?

Background

Fake news, known as yellow journalism[2] in the 1890s, has persisted for centuries, but has only recently become a household word. The internet has amplified the spread of such news and, therefore, fake news has become a mainstay of 21st century life. The World Economic Forum ranked digital wildfires, their term for the spread of misinformation and fake news, as a top global risk in 2018.[3] The motivations to spread fake news vary, but most creators seek financial gain, to advance a personal agenda, or damage an entity's reputation or brand. According to the

[1] Cavazos, Robert. 2019. The economic cost of bad actors on the internet. https://s3.amazonaws.com/media.mediapost.com/uploads/EconomicCostOfFakeNews.pdf
[2] https://www.cits.ucsb.edu/fake-news/brief-history
[3] http://www3.weforum.org/docs/WEF_GRR18_Report.pdf

Washington Post, one example occurred in August 2017 when an Internet troll “cooked up a campaign against Starbucks, posting bogus tweets that advertised ‘Dreamer Day’, when the coffee chain would supposedly give out free drinks to undocumented immigrants”[4] . With immigration being such a divisive issue in the US today, a story like this could tarnish the Starbucks brand for many consumers.

In addition, fake news “profiteers” seek financial gain by spreading false information. There are various examples in which the Securities and Exchange Commission (SEC) filed security fraud charges against people whose false information caused sharp changes in stock prices. The previously cited Washington Post article posits “fake news releases and social media posts threaten to be all the more influential on markets now that many investment firms rely on algorithms to scour press releases to inform trading decisions”.

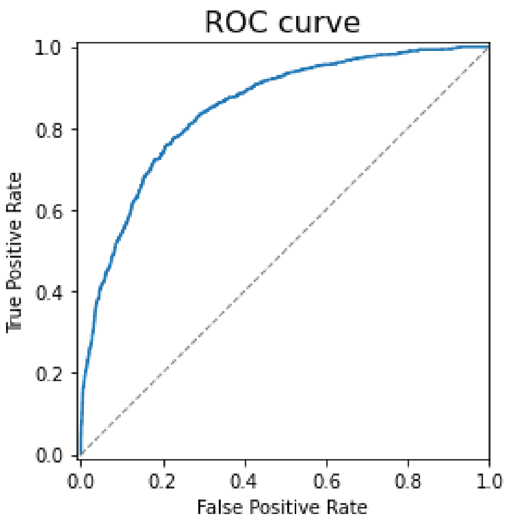
To help businesses tackle the tremendous task of mitigating the harm of misinformation to their brands, we explored the possibility of using data and analytics to help firms find and debunk fake news faster. We built a predictive model that classifies a publication as fake or real based on its headline. This tool, used with the combination of web scraping, provides businesses the ability to prevent the spread of fake news by identifying false information before it can circulate widely. By limiting the spread of misinformation, companies can protect their brands, potentially mitigate financial loss, and in limiting the propagation of false information, perhaps even dissuade bad actors from creating and disseminating fake news in the first place.

[4] https://www.washingtonpost.com/outlook/fake-news-threatens-our-businesses-not-just-our-politics/2019

Data

To build the classification model, approximately 15,000 headlines from November 2018 through November 2020 were scraped from the Reddit subreddits r/NotTheOnion and r/TheOnion via the Pushshift Reddit API[5,6]. Of these headlines, approximately 5,000 came from the popular satirical publication *The Onion* while the remaining headlines came from real news articles. While the classification model aims to identify fake news from real news, the use of these data may present challenges in generalizing this model to non-*Onion* or non-satirical sources of fake news. However, cleanly labeled "fake news" datasets are difficult to obtain. This approach provided us with the highest quality labeled data set readily available for us to build the classification model. For that reason, we consider this model to be more of a proof of concept rather than a generalizable product.

Methodology & Model Performance



To begin, we used a "countvectorizer" to count the appearance of words in the headlines and create a matrix of words to train the model. The Naive Bayes algorithm was applied to this training dataset to build the model. This approach calculates the probability of a data point belonging to the real or fake news class and then assigns each headline the class with higher probability. Model performance is described by a few metrics like AUC, Precision, and Recall. AUC is the area under the ROC curve, shown by the blue line on the chart to the right. For our model, the AUC is 0.763, which is much better than a random classifier, which has an AUC of 0.5 and is shown as the dotted grey line on the chart. In addition, the model achieved a $\text{Precision}_{\text{Onion}}$ of 0.678 and $\text{Recall}_{\text{Onion}}$ of 0.692. Precision is the rate of correctly predicted fake news classifications, while Recall is the rate of correct predictions among all fake headlines.

5 https://github.com/pushshift/api
6 https://github.com/lukefeilberg/onion/blob/master/Onion.ipynb

Following training and validation, the model was also tested on an orthogonal Kaggle dataset that contained real news headlines from Reuters and fake news headlines identified by Politifact from multiple sources. On these data, the model achieved a $\text{Recall}_{\text{Fake}}$ of 0.564. Interestingly, the model also achieved a $\text{Precision}_{\text{Fake}}$ of 0.710, which is higher than precision in the validation data. This means that the model classified few “false positives” in the Kaggle dataset.

Findings

In general, the most important features that predicted whether or not an article was fake or real can be loosely categorized, as shown in the figure to the right. The features fall into media (in green), politics (in yellow), trends (in purple), and miscellaneous (in grey).

Outcomes & Limitations

Overall, a predictive model such as the one presented here has the potential to help businesses more quickly and efficiently identify and debunk misinformation that could harm their reputation and brand. In addition, widespread use of a tool such as this could limit the ability of false information to propagate widely, and therefore, could disincentivize bad actors from creating fake news altogether.

However, using a predictive model like this to classify fake news has limitations. For example, false information can appear very different from source to source. For that reason, our model, trained using *The Onion* headlines, did not perform as well on an external dataset. A generalizable model should be trained on a diverse, high-quality labeled dataset, which may be difficult to obtain. Furthermore, since news content changes so quickly, a fake news classifier would need to be retrained frequently to continue making relevant predictions. Despite these limitations, predictive analytics offers companies clear value in the battle to protect their brands from misinformation.