

ISOT Fake News Dataset

The dataset contains two types of articles fake and real News. This dataset was collected from real-world sources; the truthful articles were obtained by crawling articles from Reuters.com (News website). As for the fake news articles, they were collected from different sources. The fake news articles were collected from unreliable websites that were flagged by Politifact (a fact-checking organization in the USA) and Wikipedia. The dataset contains different types of articles on different topics, however, the majority of articles focus on political and World news topics.

The dataset consists of two CSV files. The first file named “True.csv” contains more than 12,600 articles from reuter.com. The second file named “Fake.csv” contains more than 12,600 articles from different fake news outlet resources. Each article contains the following information: article title, text, type and the date the article was published on. To match the fake news data collected for kaggle.com, we focused mostly on collecting articles from 2016 to 2017. The data collected were cleaned and processed, however, the punctuations and mistakes that existed in the fake news were kept in the text.

The following table gives a breakdown of the categories and number of articles per category.

News	Size (Number of articles)	Subjects	
		Type	Articles size
Real-News	21417	<i>World-News</i>	<i>10145</i>
		<i>Politics-News</i>	<i>11272</i>
		Fake-News	23481
Fake-News	23481	<i>Government-News</i>	<i>1570</i>
		<i>Middle-east</i>	<i>778</i>
		<i>US News</i>	<i>783</i>
		<i>left-news</i>	<i>4459</i>
		<i>politics</i>	<i>6841</i>
		<i>News</i>	<i>9050</i>

To cite this dataset use:

1. **Ahmed H, Traore I, Saad S. "Detecting opinion spams and fake news using text classification", Journal of Security and Privacy, Volume 1, Issue 1, Wiley, January/February 2018.**
2. **Ahmed H, Traore I, Saad S. (2017) "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. In: Traore I., Woungang I., Awad A. (eds) Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments. ISDDC 2017. Lecture Notes in Computer Science, vol 10618. Springer, Cham (pp. 127-138).**