

1) Sources

Name of the source	URL	Quality of source	Number of units	Explanation of the quality of source
News Webpage	https://www.theguardian.com/world/2016/sep/18/dozens-dead-and-missing-after-typhoon-lashes-eastern-china	High	6	The quality of source is a good one since it talks about typhoon hitting the east China.
News Webpage	http://www.worldvision.org/disaster-response-news-stories/worst-natural-disasters-2015	High	29	Good one since it talks about several recent natural disaster happen and it comes from a news web page.
News Webpage	https://www.theguardian.com/australia-news/2016/sep/18/victorian-floodwaters-recede-and-great-ocean-road-reopens-after-landslides	High	11	It is from a news web page and talk about a specific natural disaster.
A Webpage by certain organization	http://gndr.org/programmes/advocacy/365-disasters/more-than-365-disasters-blogs/item/1449-types-of-recent-disasters-in-afghanistan.html	Low	10	Not a very great source since it is some viewpoint from a organization point of view.
Blog	http://blogs.forrester.com/stephanie_balaouras/15-08-31-the_state_of_business_continuity_we_have_a_long_way_to_go_to_achieve_true_resiliency	Low	15	Not a high quality since it is written by someone else and might missed some facts.
Blog	http://www.disasteraccountability.org/blog2/	Low	29	Pretty decent source in terms that it has lots of information but still the quality is low since it is a blog written by user and may have some facts which are not accurate.

1) Features

Name of Feature	representative example	Range of value
Token contains alphabet	"disaster" – 1 "2016" - 0	0-1
Token contains number	"disaster" – 0 "2016" - 1	0-1
Token contains special symbol	"\$100" – 1 "emergency" - 0	0-1
Token length	Typhoon – 7	Length of the token
Token a recent disaster	Token >= 2015 -> 1 Otherwise 0	0-1
Count of vowels in the token	Katrina – 3	Count of vowels
POS Tagging	NN country	All the possible tags for the word.

2) Results

Yes, the quality of source did matter. The source from good quality have better entities and complete information like the place of disaster, the time of disaster, cause, effect, date etc and hence it has more relevant data. On the other hand the data from the low quality source have data which we might not have seen before since these data are from blogs and can have random data based on user. To conclude the data from good quality of source have better depth of knowledge compared to low quality data.

Precision	Recall	F1	Category	Dataset (Only showing few relevant ones.)
0.5625 0.7647 0.5000 1.0000 1.0000 0.347928	0.8182 0.9750 0.2000 0.2667 0.2857 0.231415	0.6667 0.8571 0.2857 0.4211 0.4444 0.243184	10 high quality sentences	Number Irrelevant Location Name Date Macro-Average
0.6923 0.8410 0.0833 0.146970	0.7500 0.9820 0.1667 0.172609	0.7200 0.9061 0.1111 0.157926	10 low quality sentences	Number Irrelevant Location Macro-Average
0.6207 0.8075 0.1875 1.0000 1.0000 0.328696	0.7826 0.9791 0.1875 0.2222 0.1538 0.211388	0.6923 0.8850 0.1875 0.3636 0.2667 0.217741	for all 20 sentences	Number Irrelevant Location Name Date Macro-Average