COLLECTION OF DATA SETS

TEAM NO	PNT2022TMID13948
PROJECT NAME	Project – Natural Disaster Intensity Analysis and Classification Using Artificial Intelligence

Feature	Explanation
Brown Cluster Ngrams	Unigram and Bigram of 1000 Brown clusters in Twitter Word Clusters made available by CMU ARK group
Count of disaster related terms	Number of disaster related terms as informative words in a created lexicon for disaster tweets.
Total PMI Score of disaster related terms	Total PMI scores of unigrams and bigrams words that occurred in the tweet and listed as strongly correlated with natural disaster in a disaster lexicon for tweets
Count of non-informational terms	Lexicon creation function of this system also identifies a set of terms which appear only in Not-Informative tweets across all natural disasters datasets.
Total PMI Score of non-informational terms	Total PMI Score for each set of unigrams, bigrams that mostly occur in the Not informative tweet.
Count of numerals	Expected to be higher in situational tweets which contain information such as the number of casualties, emergency contact numbers.
POS tag	Unigram part of speech tags that occur in the Tweet generated by CMU ARK POS-Tagger
Word Ngrams	Unigram and bigram of terms from Disaster Lexicon

■ Open Data	21	■ Restricted	16	■ Closed	0	■ Unknown	17
Free to access, use and share		Technical, legal or cost restrictions		Access not permitted or does not exist		No information. Submission is needed	

Show: Submitted, Not submitted, All

Dataset	Open data criteria
Active seismic fault GEM Global Active Faults Global Earthquake Model (GEM) Foundatio 2018	⊘ 🖵 ☁ 🏷 ⧉ ⌨ 👁 $ 🔓 🕐
Flood protection measures FLOPROS: an evolving global database of flood protection standards 2016	⊘ 🖵 ☁ 🏷 ⧉ ⌨ 👁 $ 🔓 🕐
Land cover MODIS Land Cover Global Land Cover Facility, University of Maryland 2014-09-23	⊘ 🖵 ☁ 🏷 ⧉ ⌨ 👁 $ 🔓 🕐
Nearshore tsunami wave height Global Tsunami Model - global hazard map (maximum inundation heig 2017	⊘ 🖵 ☁ 🏷 ⧉ ⌨ 👁 $ 🔓 🕐
Risk indicators from previous studies Risk indicators from previous studies World Resource Institute March 2015	⊘ 🖵 ☁ 🏷 ⧉ ⌨ 👁 $ 🔓 🕐
Site conditions map Global Vs30 Mosaic United States Geological Survey (USGS) Unsure	⊘ 🖵 ☁ 🏷 ⧉ ⌨ 👁 $ 🔓 🕐
Volcanoes Global Volcanism Program database Smithsonian Institution December 21, 2018	⊘ 🖵 ☁ 🏷 ⧉ ⌨ 👁 $ 🔓 🕐

Water bodies

Type	Disaster Name	Info	Not-Info	Total
Floods	2013_Queensland_floods (QF)	728	281	1009
Bushfire	2013_Australia_bushfire (AB)	691	261	952
Typhoon	2013_Typhoon_Yolanda (TY)	765	175	940
Wildfire	2012_Colorado_wildfires (CW)	685	247	932
Earthquake	2014_Chile_earthquake (ChiE)	1834	179	2013
Floods	2013_Colorado_floods (CF)	589	190	779
Earthquake	2014_Costa_Rica_earthquake (CE)	842	170	912
Floods	2014_Manila_floods (MF)	628	293	921
Floods	2012_Phillipines_Floods (PF)	761	145	906
Floods	2013_Alberta Floods (AF)	684	297	981
Floods	2014_India_floods (IF)	940	396	1336

Dataset	Type 1	Type 2	Type 3	Type 4	Type 5
QF	207	113	55	114	17
AB	199	65	35	70	33
TY	77	106	383	20	63
CW	44	128	62	69	25
NE	6	165	239	1215	458
IF	30	792	42	51	25
ChiE	55	3	70	14	58

Dataset	Precision	Recall	F1	Accuracy
AB	0.897	0.895	0.895	89.45%
TY	0,912	0,92	0,913	92,02%
IF	0.908	0.908	0.908	90.82%
NE	0.748	0.751	0.749	75.05%