Richter's Predictor: Modeling Earthquake Damage

Alperen KANTARCI
ITU 2018-2019 Machine Learning Term Project

Ranking and Leaderboard

BEST	CURRENT RANK	# COMPETITORS	SUBS. MADE
0.7490	39	1573	0 of 3

chth	37	0.7491	2019-11-09 16:35:13	10
ZHANGJUNRUI666	38	0.7490	2019-10-07 00:32:48	23
Alpkant	39	0.7490	2019-12-20 13:36:26	13
manhitv	40	0.7489	2019-11-20 09:45:28	12
mglowacki100@gmail.com	41	0.7488	2019-08-03 06:09:44	1
hamonk2	42	0.7488	2019-11-17 21:46:24	6
lucasd	43	0.7488	2019-11-13 09:52:15	3
arpan65	44	0.7487	2019-12-22 10:40:51	4
lhore	45	0.7486	2019-12-23 11:14:44	16
Daniel_Terol_UGR	46	0.7486	2019-12-24 13:26:21	40
gobearx	47	0.7486	2019-10-14 06:32:05	1
rio2020	48	0.7485	2019-12-22 09:00:44	5
mohawker	49	0.7485	2019-11-04 17:03:02	10
mosiomohsen	50	0.7484	2019-06-22 09:56:06	6
Benchmark: Random Forest		0.5815		

Submission and models

1-) **Naïve Bayes** with dropped features (as a baseline performance): 0.5642

2-) **Decision Tree** with dropped features and random parameters : 0.4455

3-) One hot encoded and normalized features with **random forest**: 0.7139

4-) **Gradient boosting** with new features and label encodings : 0.7490

SUBMISSIONS

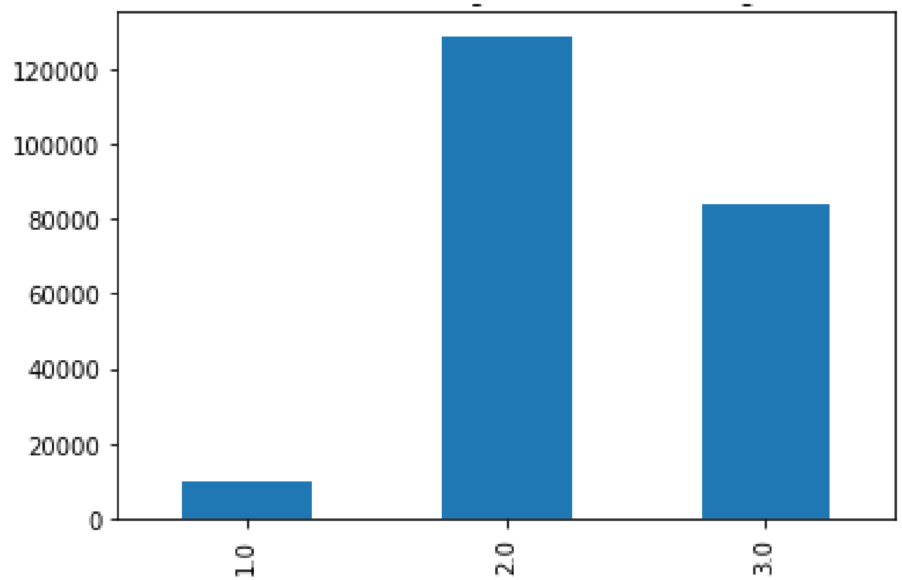
Score	Submitted by	Timestamp
0.5642	Alpkant	2019-11-19 12:26:57 UTC
!	Alpkant	2019-11-19 12:40:41 UTC
0.4455	Alpkant	2019-11-19 12:41:21 UTC
0.4439	Alpkant	2019-11-19 12:42:00 UTC
0.7139	Alpkant	2019-11-20 04:46:35 UTC
0.7120	Alpkant	2019-11-20 08:58:07 UTC
0.5670	Alpkant	2019-11-21 10:13:46 UTC
0.7400	Alpkant	2019-12-19 19:49:23 UTC
!	Alpkant	2019-12-19 19:53:51 UTC
0.7443	Alpkant	2019-12-20 12:46:52 UTC
0.7443	Alpkant	2019-12-20 13:20:47 UTC
0.7490	Alpkant	2019-12-20 13:36:26 UTC
0.7476	Alpkant	2019-12-21 05:57:01 UTC
0.7463	Alpkant	2019-12-21 09:56:20 UTC

Features

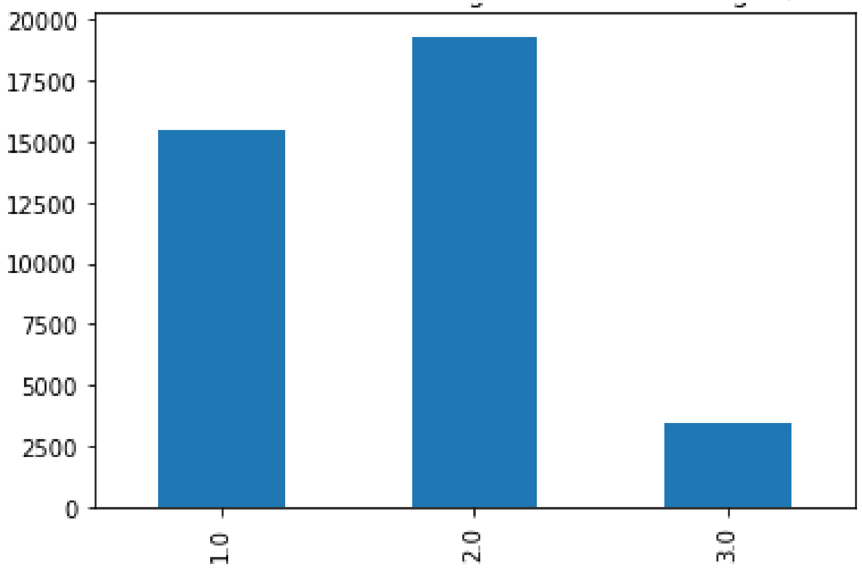
1-) Building material

Super structures can be grouped into 4: mud, cement, concrete and wooden. There can be correlation with the damage grade. We see that mud buildings are damaged more than non-mud buildings

Mud buildings



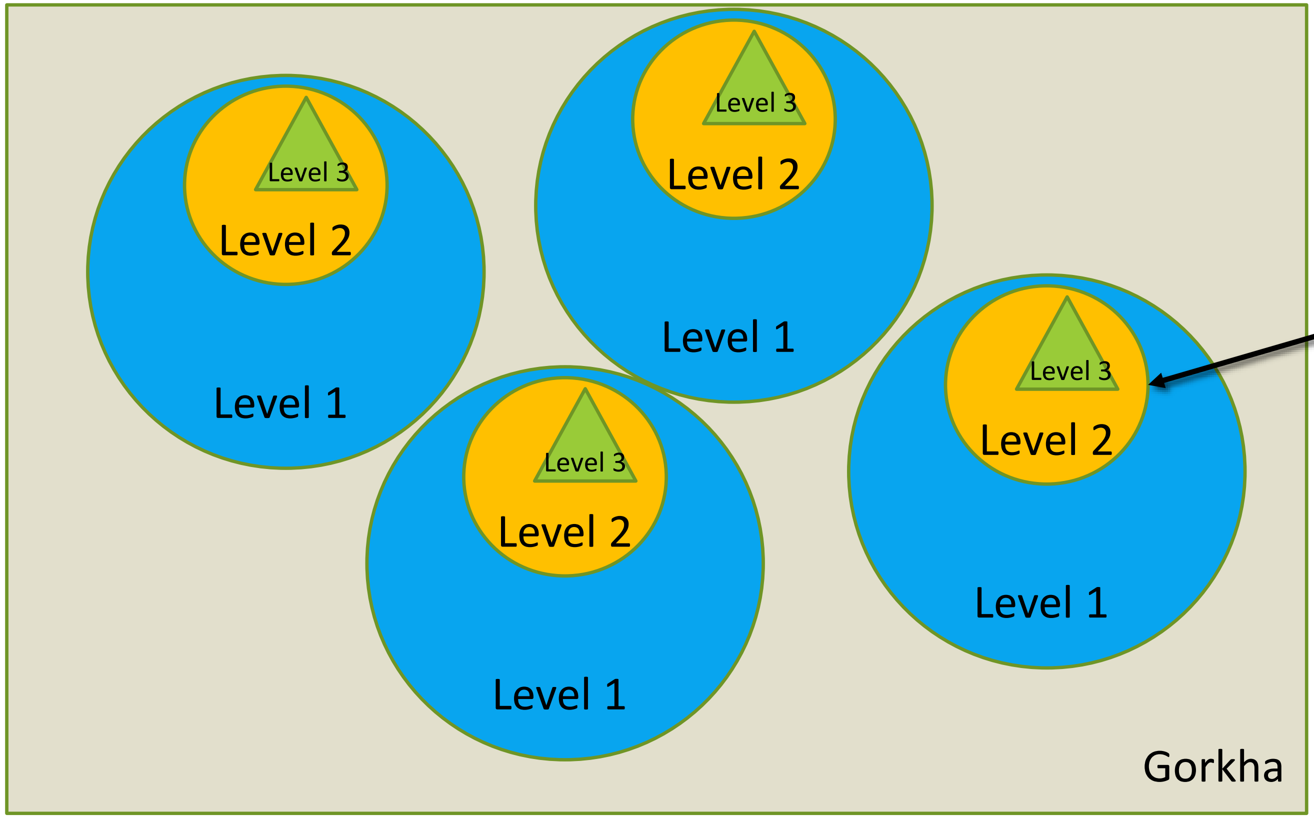
Cement – concrete - wooden buildings



Features

2-) Geographic location

For each level, damage grades of the mud, concrete, cement and wooden buildings are calculated. Both building and geo-locational damage features are created. Some location ids may not be in both sets.



More mud buildings ?
Or close to the center of the earthquake ?

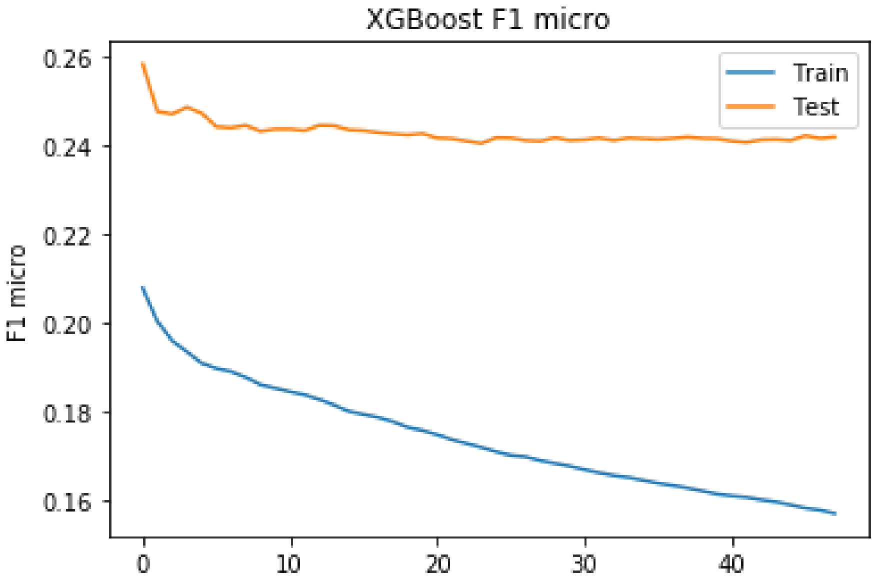
Correlation map after new features

Features and training

3-) 8 categorical data converted by using Frequency Encoder. Frequencies of the categories converted to a feature

Gradient Boosting Machine build trees one at a time, where each new tree helps to correct errors made by previously trained tree. It needs careful parameter tuning. Exhaustive parameter search could increase the accuracy to 0.75 but hand selected parameters have been tried.

Validation set f1 score: 0.7595



	precision	recall	f1-score	support
1.0	0.70	0.55	0.61	1232
2.0	0.76	0.86	0.80	7432
3.0	0.78	0.65	0.71	4367
accuracy			0.76	13031
macro avg	0.74	0.69	0.71	13031
weighted avg	0.76	0.76	0.76	13031