Predicting earthquake building damage

Amy Kim

18th May 2022

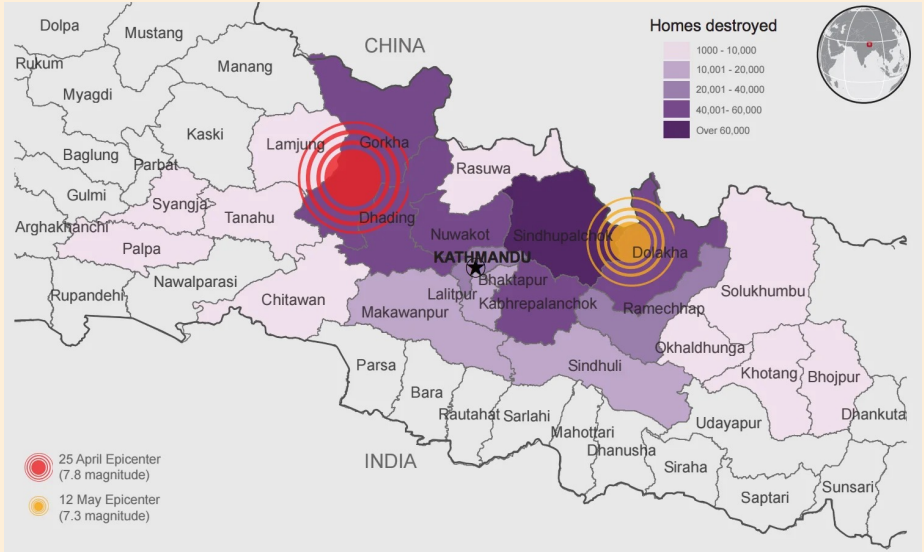
Motivation

In 2015 a 7.8 Mw Gorkha Earthquake occurred on April 25, 2015 in Nepal

- Economic cost = **$10 billion** nearly half of its GDP of $19billion
- **9,000 lives lost**

Goal:

Use machine learning classification to predict building damage



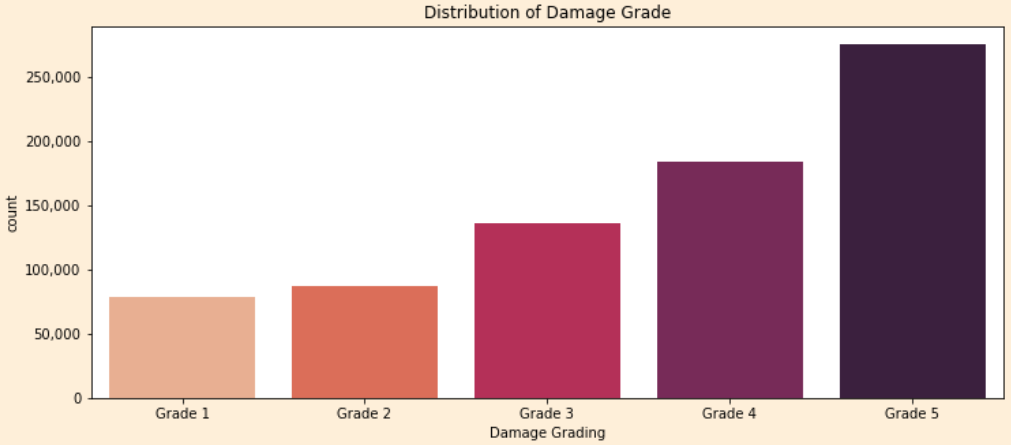
Data

Nepal carried out a household survey for 11 severely affected districts to assess building damage.

- 750k +rows , 42 features

- Target: Damage Grade
- Features: Building data

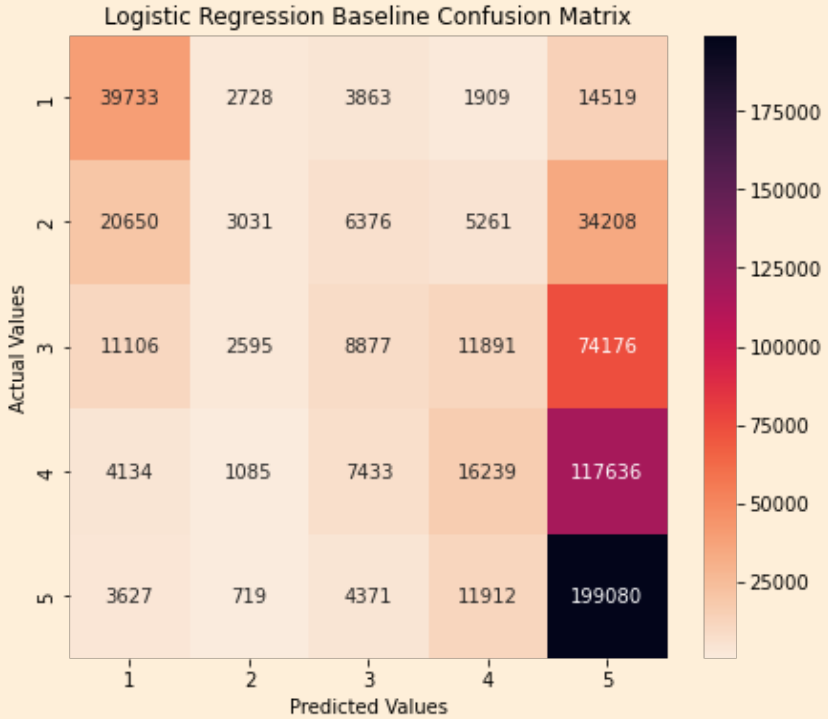
  Examples:
  - Year built
  - Construction material
  - Building height & floor area
  - Land surface



Grade 1	Thin cracks in plaster
Grade 2	Cracks in many walls
Grade 3	Large and extensive cracks
Grade 4	Walls collapse, failure of roof/floor
Grade 5	Total or near collapse

Baseline model

Logistic Regression



Log loss score: 1.32

Should we consider Binary model?

Multiclass

Class	Description	Low	High	Midpoint
Grade 1	no need/some minor	-	1,000	**500**
Grade 2	minor repair	1,000	2,500	**1,750**
Grade 3	major repair	2,500	10,000	**6,250**
Grade 4	reconstruction/some	10,000	25,000	**17,500**
Grade 5	reconstruction	25,000	50,000	**37,500**

Class	Count of Predicted	Predicted Cost	Count of Actual	Actual Cost
Grade 1	79,250	$40m	62,752	$31m
Grade 2	10,158	$18m	69,526	$122m
Grade 3	30,929	$193m	108,645	$679m
Grade 4	47,212	$826m	146,536	$2,564m
Grade 5	439,619	$16,486m	219,709	$8,239m
	607,168	**$17,563m**	**607,168**	**$11,636m**

Model miss: $6 billion

Binary

Class	Description	Low	High	Midpoint
0	No need/Minor repair	-	2,500	**1,250**
1	Major repair/reconstruction	5,000	50,000	**27,500**

Class	Count of Predicted	Predicted Cost	Count of Actual	Actual Cost
0	89,408	$112m	132,278	$66m
1	517,760	$14,238m	474,890	$831m
	607,168	**$14,350m**	**607,168**	**$897m**

Model miss:

$13 billion

Comparison stage 1

Classifiers

	Log loss score
Logistic Regression (baseline)	1.32
Logistic Regression (One v Rest)	1.31
Random Forest Classifier (SMOTE)	1.24
XG Boost Classifier (SMOTE)	1.27

Regressors

	RMSE
Linear Regression (baseline)	1.12
Random Forest Regressor (SMOTE)	0.99
XG Boost Regressor (SMOTE)	1.02

Comparison stage 2

Classifiers

	F1 score (micro)
Random Forest Classifier (SMOTE)	0.54

Regressors

	F1 score (micro)
Random Forest Regressor (SMOTE)	0.43

Ensemble

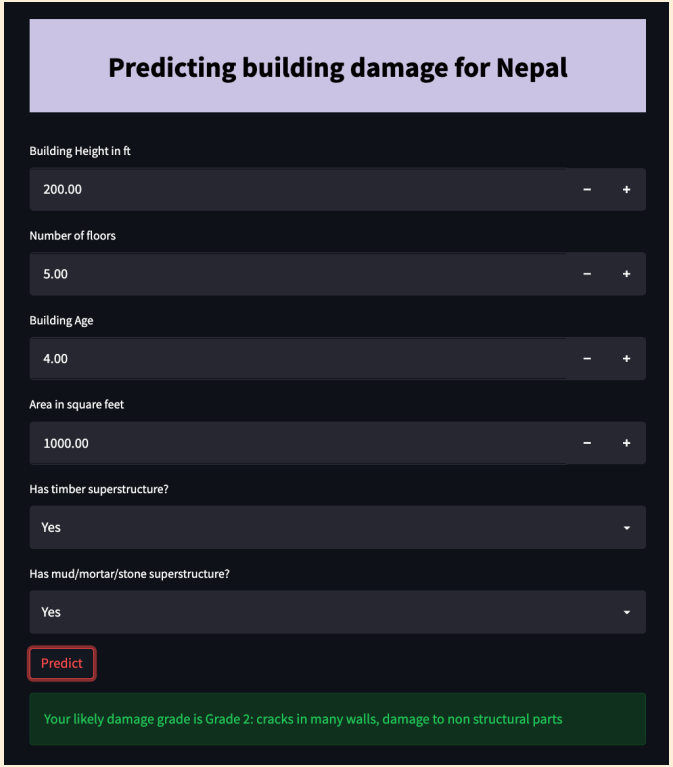
	F1 score (micro)
RF Regressor (SMOTE) RF Classifier (SMOTE) XGB Regressors (SMOTE) XGB Classifier (SMOTE)	0.45

Random Forest Classifier F1 score on final holdout : 0.45

Model Deployment

Top 6 features were used to deploy a simple proof of concept model

Please visit at : https://share.streamlit.io/amyyunekim/course_4_classification/main/app/app.py



Future development

- Further tuning of hyperparameters
- Use cloud computing for faster processing on large data sets

Discussion

- Overfitting problem training score was higher than test score.
  - CV only included 3 folds
  - Use cloud computing for faster processing on large data sets

- Further tuning of hyperparameters

Thanks!

Do you have any questions?

amyyunekim@gmail.com