**PRCP-1015-EquakeDamagePred**

**Problem Statement**

Task 1:-Prepare a complete data analysis report on the given data.

Task 2:-Create a predictive model in such a way that the objective is to predict the ordinal variable “damage\_grade”. This column presents the level of damage grade affected by the earthquake.

Task3:-Suggestions to the Seismologists to avoid earthquake event leading to significant damage in many buildings

**Dataset Link:**

You will be predicting the amount of damage buildings receive during an earthquake. This is a [Driven Data competition](https://www.drivendata.org/competitions/57/nepal-earthquake/).

We’ve set up this Kaggle classroom competition so multiple teams within MDST can compete. The winning model will be used for our final submission to the Driven Data leaderboard.

From the Driven Data website:

We're trying to predict the ordinal variable damage\_grade, which represents a level of damage to the building that was hit by the earthquake. There are 3 grades of the damage:

1 represents low damage

2 represents a medium amount of damage

3 represents almost complete destruction

**Link:** <https://d3ilbtxij3aepc.cloudfront.net/projects/CDS-Capstone-Projects/PRCP-1015-EquakeDamagePred.zip>

## **Features**

* The dataset mainly consists of information on the buildings' structure and their legal ownership. Each row in the dataset represents a specific building in the region that was hit by Gorkha earthquake.
* There are 39 columns in this dataset, where the building\_id column is a unique and random identifier. The remaining 38 features are described in the section below. Categorical variables have been obfuscated random lowercase ascii characters. The appearance of the same character in distinct columns does not imply the same original value.

## **Description**

* · geo\_level\_1\_id, geo\_level\_2\_id, geo\_level\_3\_id (type: int): geographic region in which building exists, from largest (level 1) to most specific sub-region (level 3). Possible values: level 1: 0-30, level 2: 0-1427, level 3: 0-12567.
* · count\_floors\_pre\_eq (type: int): number of floors in the building before the earthquake.
* · age (type: int): age of the building in years.
* · area\_percentage (type: int): normalized area of the building footprint.
* · height\_percentage (type: int): normalized height of the building footprint.
* · land\_surface\_condition (type: categorical): surface condition of the land where the building was built. Possible values: n, o, t.
* · foundation\_type (type: categorical): type of foundation used while building. Possible values: h, i, r, u, w.
* · roof\_type (type: categorical): type of roof used while building. Possible values: n, q, x.
* · ground\_floor\_type (type: categorical): type of the ground floor. Possible values: f, m, v, x, z.
* · other\_floor\_type (type: categorical): type of constructions used in higher than the ground floors (except of roof). Possible values: j, q, s, x.
* · position (type: categorical): position of the building. Possible values: j, o, s, t.
* · plan\_configuration (type: categorical): building plan configuration. Possible values: a, c, d, f, m, n, o, q, s, u.
* · has\_superstructure\_adobe\_mud (type: binary): flag variable that indicates if the superstructure was made of Adobe/Mud.
* · has\_superstructure\_mud\_mortar\_stone (type: binary): flag variable that indicates if the superstructure was made of Mud Mortar - Stone.
* · has\_superstructure\_stone\_flag (type: binary): flag variable that indicates if the superstructure was made of Stone.
* · has\_superstructure\_cement\_mortar\_stone (type: binary): flag variable that indicates if the superstructure was made of Cement Mortar - Stone.
* · has\_superstructure\_mud\_mortar\_brick (type: binary): flag variable that indicates if the superstructure was made of Mud Mortar - Brick.
* · has\_superstructure\_cement\_mortar\_brick (type: binary): flag variable that indicates if the superstructure was made of Cement Mortar - Brick.
* · has\_superstructure\_timber (type: binary): flag variable that indicates if the superstructure was made of Timber.
* · has\_superstructure\_bamboo (type: binary): flag variable that indicates if the superstructure was made of Bamboo.
* · has\_superstructure\_rc\_non\_engineered (type: binary): flag variable that indicates if the superstructure was made of non-engineered reinforced concrete.
* · has\_superstructure\_rc\_engineered (type: binary): flag variable that indicates if the superstructure was made of engineered reinforced concrete.
* · has\_superstructure\_other (type: binary): flag variable that indicates if the superstructure was made of any other material.
* · legal\_ownership\_status (type: categorical): legal ownership status of the land where building was built. Possible values: a, r, v, w.
* · count\_families (type: int): number of families that live in the building.
* · has\_secondary\_use (type: binary): flag variable that indicates if the building was used for any secondary purpose.
* · has\_secondary\_use\_agriculture (type: binary): flag variable that indicates if the building was used for agricultural purposes.
* · has\_secondary\_use\_hotel (type: binary): flag variable that indicates if the building was used as a hotel.
* · has\_secondary\_use\_rental (type: binary): flag variable that indicates if the building was used for rental purposes.
* · has\_secondary\_use\_institution (type: binary): flag variable that indicates if the building was used as a location of any institution.
* · has\_secondary\_use\_school (type: binary): flag variable that indicates if the building was used as a school.
* · has\_secondary\_use\_industry (type: binary): flag variable that indicates if the building was used for industrial purposes.
* · has\_secondary\_use\_health\_post (type: binary): flag variable that indicates if the building was used as a health post.
* · has\_secondary\_use\_gov\_office (type: binary): flag variable that indicates if the building was used fas a government office.
* · has\_secondary\_use\_use\_police (type: binary): flag variable that indicates if the building was used as a police station.
* · has\_secondary\_use\_other (type: binary): flag variable that indicates if the building was secondarily used for other purposes.

**Model Comparison Report**

Create a report stating the performance of multiple models on this data and suggest the best model for production.

**Report on Challenges faced**

Create a report which should include challenges you faced on data and what technique used with proper reason.

Note:-All above task has been created on single jupyter notebook and share the same while final submission of project.