***Post earthquake bridge damage prediction using machine learning algorithms***

Yash Garg

B. Tech CSE

ASET, Amity University

Noida, India

[ygarg704@gmail.com](mailto:skhiearth@gmail.com)

Arpit Masih

B. Tech CSE

ASET, Amity University

Noida, India

glitcheritzu@gmail.com

*Abstract*— More the magnitude of the earthquake, more is the damage. During an earthquake, the mobility of the soft soil gets augmented and because most of the bridges are built on soft soil, there are even more chances of the bridge behaving like a ship in the sea. Stability pf the bridges is the most crucial task to avoid all disasters. Many bridges collapse because their mobility and sustainability cannot stand the magnitude of the earthquake. In this paper, we have proposed a method to predict the damages that a bridge will sustain after an earthquake, using many 3 classification algorithms like k-nearest neighbors, decision tree and random forest classification. This prediction in turn will help to improve bridge sustainability during an earthquake, which can help save many lives.

*Keywords*— *Earthquake, Classification, Bridges, Sustainability, Machine learning, Magnitude.*

# **Introduction**

The condition of bridges has always been a concern for the engineers as well as for the public. With the increase in age and degradation of materials used, an earthquake of a certain level might be catastrophic. The measures of safety have changed over the years, and can depend from case to case, and with the help of the current industrial technology, the construction has gotten efficient but we do see the old bridges experiencing quite some damage. Therefore, if we had an application that could help us measure, in some way, the damage a bridge undergoes post-earthquake, with the help of that, the engineers could formulate some important data from it and use it in future bridge designs.

In this report we have tried to make a model that predicts the damage level of a bridge post-earthquake. That is, how much damage a bridge might undergo if it were to be hit by an earthquake of certain amplitude. The features we have taken into consideration are :

1. Material Type
2. Structure Type
3. Age
4. Distance from the epicentre

The classification algorithms we used in this model were K Neighbours, Decision Tree and Random Forest which predict whether the bridge will be damaged or not, and also the damage level if it is the former.

# **Related work**

The 2018 study done by Ziyang and Xiao [1], greatly inspired us to make this model. Reading some papers on similar topics like in paper [1], the authors predicted the damaged sustained by a bridge using regression algorithms. In paper [2], the authors predicted the structural damage using classification algorithms.

We added a few of them to the factors affecting the prediction and were able to get a more accurate result. In paper [3], the authors evaluated the seismic damage of bridge portfolios using similar ML techniques from which we got the idea of predicting the amount of damage sustained by a bridge on a scale, if there will be any, after an earthquake.

# **Proposed Technique**

In this paper, primarily we cleaned the data by removing the missing values and unwanted rows. After that, the data was analysed and was visualized for more understanding of the data. At that point we used classification algorithms of machine learning like decision tree, random forest and more for the prediction of the damage sustained by the bridge after an earthquake.

Next, we found the feature on the basis of which we would classify our dataset. Then, after we decided to classify our dataset on the basis of whether the bridge is damaged or not and if it is damaged then what is the damaged level. The damage level was divided into 4 parts, which are ‘no damage’, ‘moderate/severe damage’, ‘partial/overall damage’ and finally ‘total collapse’.

Then we repeated the following steps for decision tree and random forest classification so that we can get a higher accuracy score.

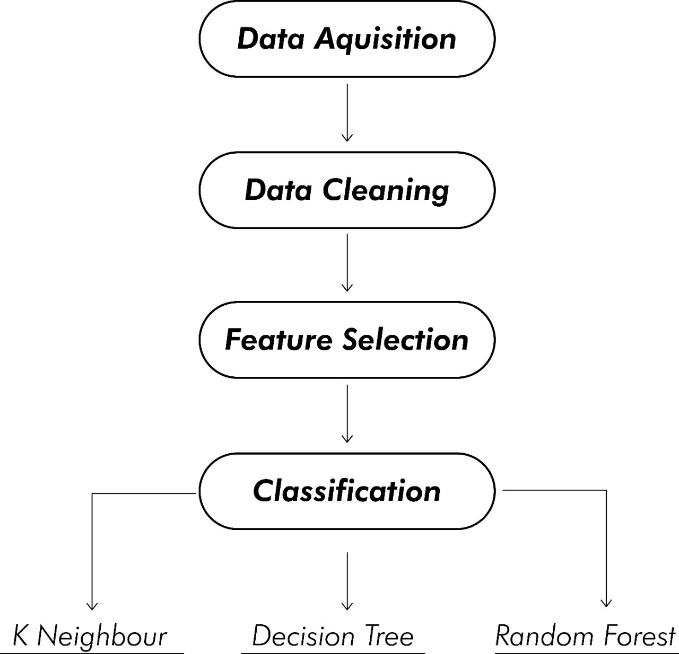


Figure 1. Application Framework.

# **Data Analysis, Visualization and Classification**

## **Data acquisition, Cleaning and Visualization**

We collected the dataset with 250 entries of “San Francisco-Oakland Bay Bridge”, “Interstate 5 (Golden State Freeway), Gavin Canyon”, “Interstate 10 (over La Cienega Boulevard)”, “Interstate 405 (over Jefferson Boulevard)” and many more bridges. The datset was taken from github and unwanted rows and columns were removed. We used many attributes for classification like material\_type of which the bridge is made of. The type I material is steel which 144 bridges use and the type II material is concrete which 105 bridges use, as shown in fig(2).

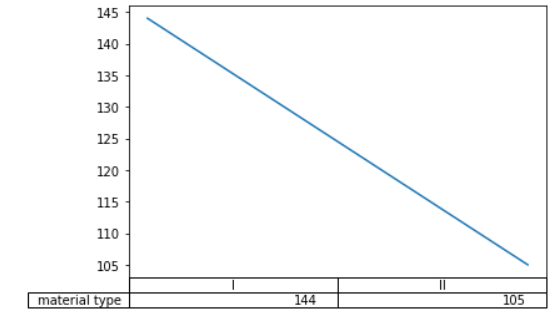


Figure 2. Material Type.

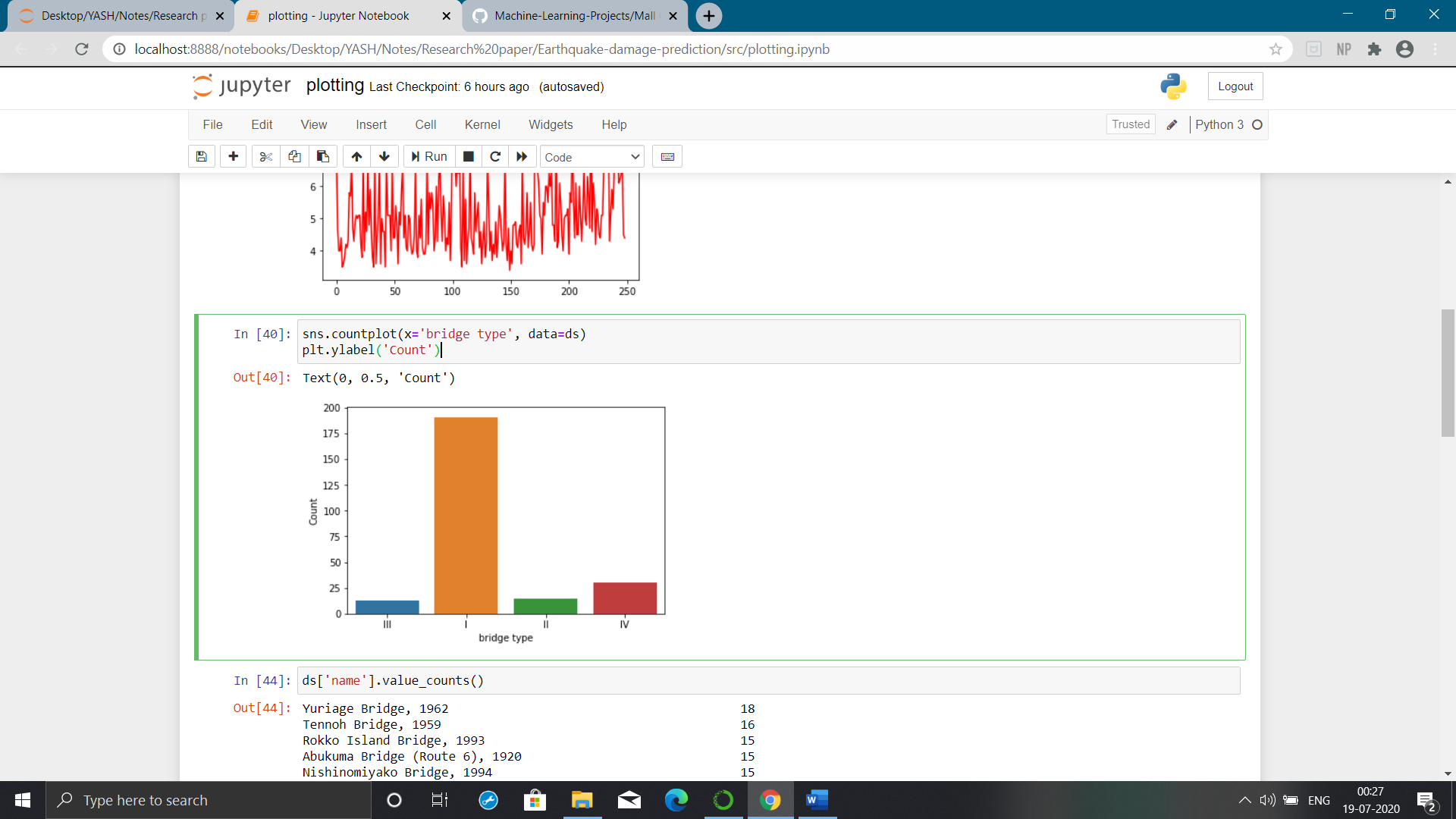


Figure 3. Bridge types.

Other features consist of the age of the bridge and the type of the bridge which is classified into 4 types which are beam (I), truss (II), cable (III), arch (IV) bridge which is shown in fig(3). It also shows that most of the bridges are of beam type. The dataset also contains a column for the epicenter of the bridge which is one of the most crucial attribute for bridge stability. We also used the magnitude of the earthquake which is the most crucial of all the attributes as shown in fig(4).

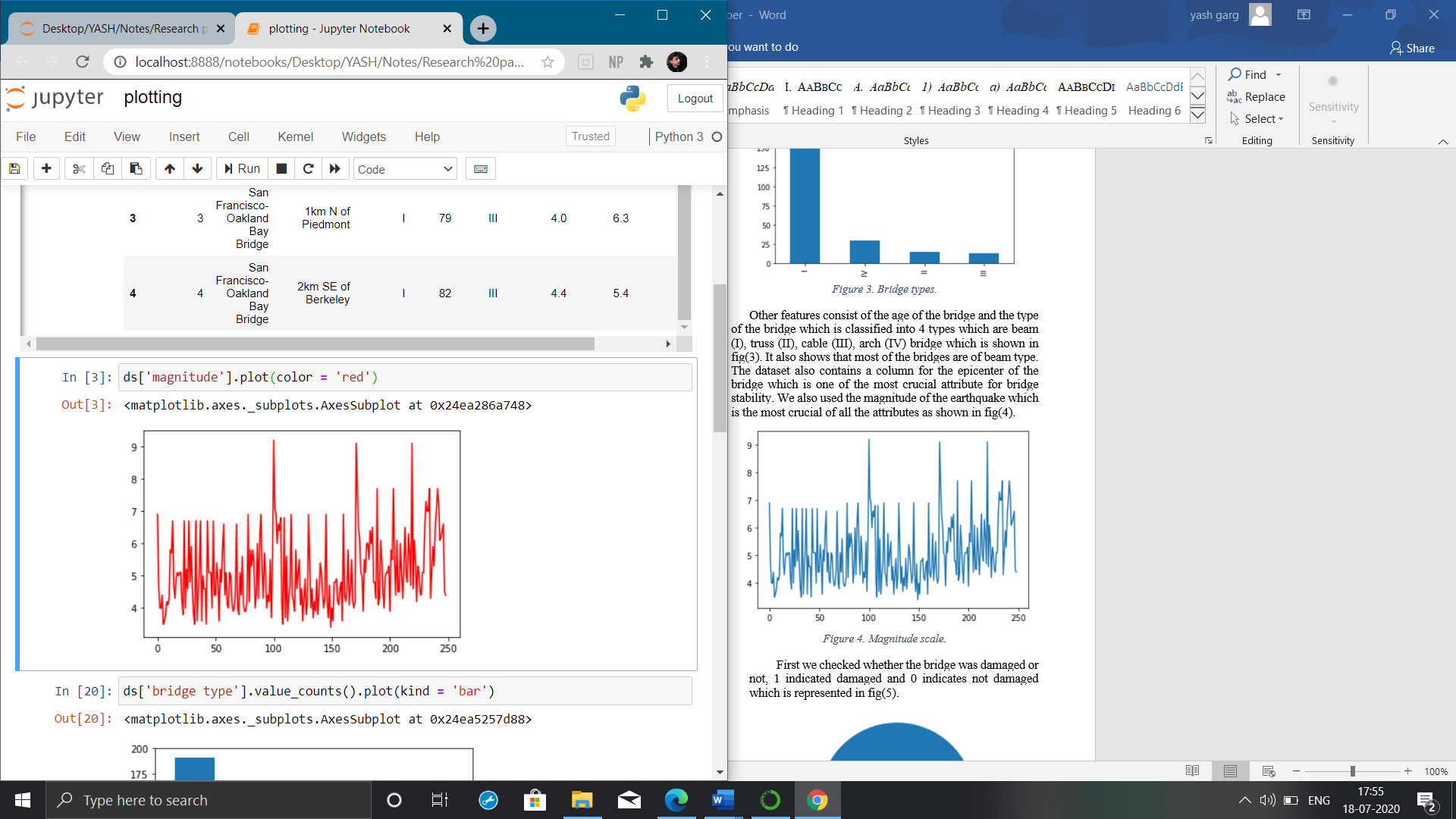


Figure 4. Magnitude scale.

First we checked whether the bridge was damaged or not, 1 indicated damaged and 0 indicates not damaged which is represented in fig(5).

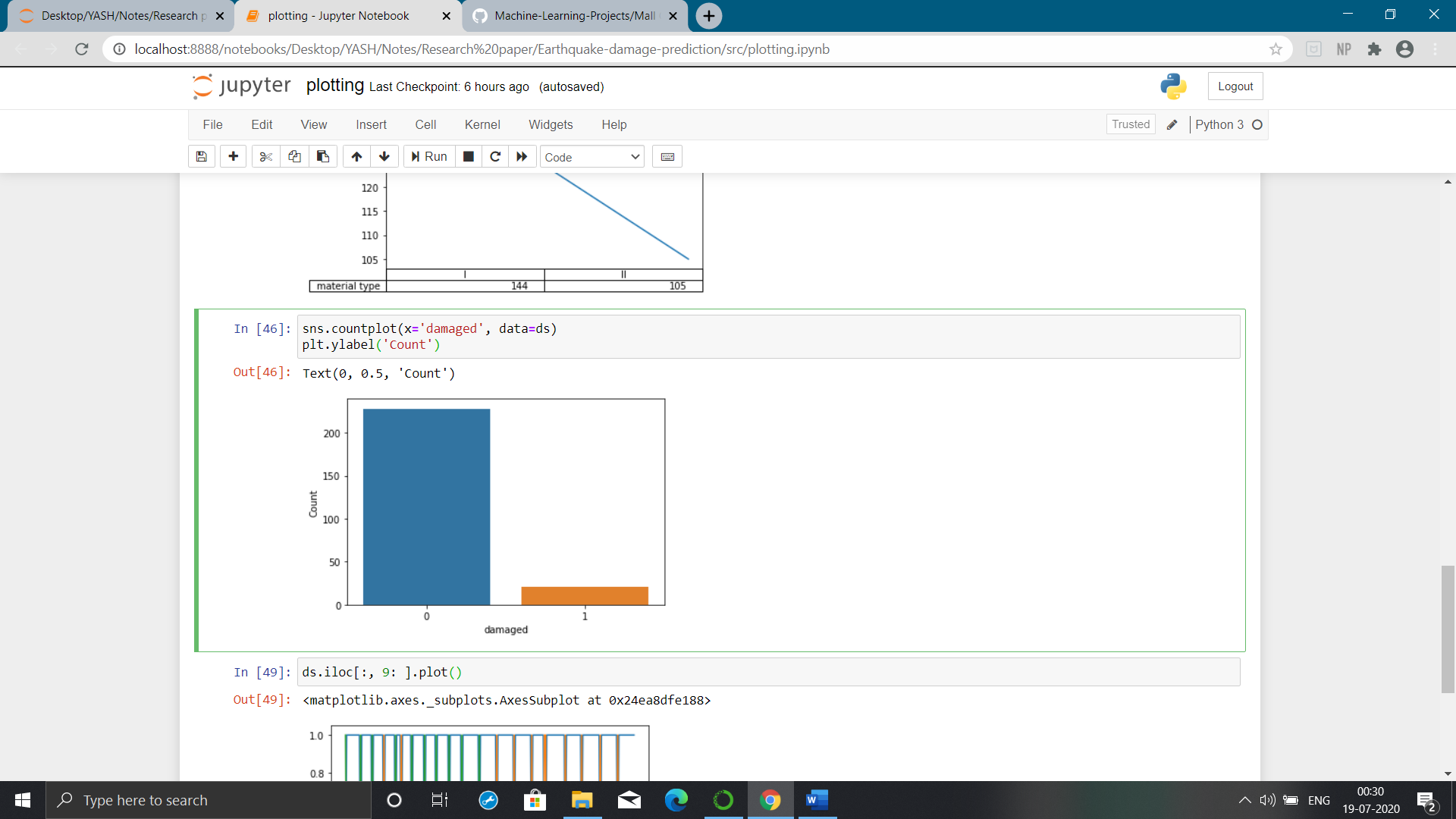


Figure 5. Damaged or not.

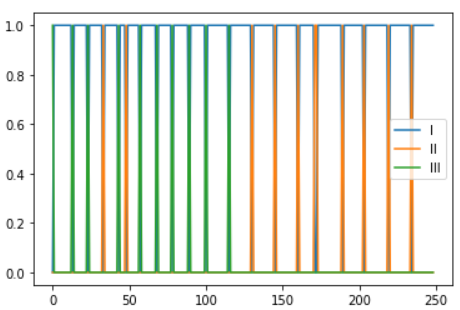


Figure 6. Damage level.

Furthermore, we used the damage level, which is divided into 4 levels which are no damage, moderate damage, partial damage and total collapse depicted in fig(6) and as we can see most of the bridges with respect to other attributes moderately damaged.

## **Classification**

After reviewing the whole dataset and plotting all the necessary graphs, we performed classification on the dataset to predict the damage sustained by the bridges. The dataset was slipt into 6:3 columns, among which the 6 columns were the part of the independent variable whereas the rest 3 columns were a part of the dependent variable. We used 3 classification methods to predict the outcomes. For the classification to the done the dataset was split into test set and training set where the split size was 0.25. The first algorithm was k-nearest neighbors, which operates by finding the distances between a query and all the selected specimens in the results, indicating the defined number specimens (K, which in this case is equal to 5) nearest to the query, then voting for the most frequent label (in the classification case) or averaging the labels (in the regression case).

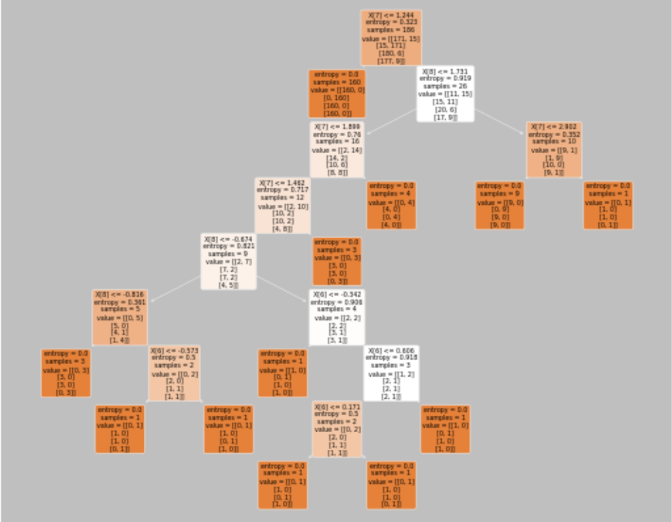


Figure 7. Decision tree plot.

Another algorithm which was taken was decision tree classification for which we got the maximum precision, f1 and recall score.

Each node consists of 4 values as we can see in fig(7), which are entropy, values, samples and X[]. The “values” entity describe the number of observations that were processed in that node and they fall into 3 categories as denoted by the color of each node.

Entopy can be described as the degree of expanse of arbitrary content in the dataset, the more arbitrary content in the dataset the more problematic it is to lure suppositions from the dataset. Another algorithm was random forest classification in which the n\_estimators were equal to 10.

## **Tools Used**

Numpy, pandas and matplotlib / seaborn packages were used for the classification of the model, cleaning of the dataset and visualization of the data.

Python language was used for the whole implementation whic was done in Jupyter Notebook. Python was used because of its high availability of libraries and packages.

# **Results**

The dataset comprises of 250 rows, each with a unique value. Score calculators like precison, f1, recall were used to calculate which classified model was the best. Precision quantifies the number of predictions of positive class that actually belong to the positive class. Recall quantifies the number of positive class predictions in the dataset made from all the positive examples. F-Measure provides a single score which balances both precision and recall concerns in one number and so we can see the decision tree classification model is the best classification model for this prediction.

|  |  |  |  |
| --- | --- | --- | --- |
| Classification algorithm | Precision | F1 | Recall |
| K-nearest neighbors | 0.80 | 0.81 | 0.84 |
| Decision tree | ***0.94*** | ***0.92*** | ***0.92*** |
| Random forest | 0.84 | 0.83 | 0.84 |

Table . Classification algorithms and their score.

## **Conclusion and future work**

Road lane detection application is very effective and efficient but it still has its disadvantages. In this paper, I presented a model of road lane detection using computer vision. The model I designed is not very effective in night images. The images should have proper lightning and should be from the down view, the image should not be from the top view or else the lane would not be detected. As shown in figure 8, the lane is not shown properly, the model did not pick the pixels correctly of the image shown

should be from the down view, the image should not be from the top view or else the lane would not be detected.

In the neat future, this application can be well used by auto driven automobiles, they can detect lanes and it will be very convenient for them and other applications is object detection. With both of these applications the total net rate of road accidents can be decreased to significantly zero.[10]

##### **REFERENCES**

1. Wang, Yue, Dinggang Shen, and Eam Khwang Teoh. "Lane detection using spline model." Pattern Recognition Letters 21.8 (2000): 677-689.
2. Sun, Zehang, George Bebis, and Ronald Miller. "On-road vehicle detection using Gabor filters and support vector machines." 2002 14th International Conference on Digital Signal Processing Proceedings. DSP 2002 (Cat. No. 02TH8628). Vol. 2. IEEE, 2002.
3. McCall, Joel C., and Mohan M. Trivedi. "An integrated, robust approach to lane marking detection and lane tracking." IEEE Intelligent Vehicles Symposium, 2004. IEEE, 2004.
4. Zhou, Yong, et al. "A robust lane detection and tracking method based on computer vision." Measurement science and technology 17.4 (2006): 736.
5. Truong, Quoc-Bao, and Byung-Ryong Lee. "New lane detection algorithm for autonomous vehicles using computer vision." 2008 International Conference on Control, Automation and Systems. IEEE, 2008.

Chiu, Kuo-Yu, and Sheng-Fuu Lin. "Lane detection using color-based segmentation." IEEE Proceedings. Intelligent Vehicles Symposium, 2005.. IEEE, 2005.

Yoo, Ju Han, et al. "A robust lane detection method based on vanishing point estimation using the relevance of line segments." IEEE Transactions on Intelligent Transportation Systems 18.12 (2017): 3254-3266.

Daigavane, Prema M., and Preeti R. Bajaj. "Road lane detection with improved canny edges using ant colony optimization." 2010 3rd International Conference on Emerging Trends in Engineering and Technology. IEEE, 2010.

Ieng, S-S., et al. "A new multi-lanes detection using multi-camera for robust vehicle location." IEEE Proceedings. Intelligent Vehicles Symposium, 2005.. IEEE, 2005.

Wang, Yue, Eam Khwang Teoh, and Dinggang Shen. "Lane detection and tracking using B-Snake." Image and Vision computing 22.4 (2004): 269-280.