Attenuation Relations between Iran and Sharjah horizontal peak ground acceleration

Abstract:

Iran is one of the most seismic countries of the world. It is situated over the Himalayan-Alpide seismic belt and is one of those countries, which have lost many human lives and a lot of money due to the occurrence of earthquakes. Here a model is built using Machine Learning to predict PGA in this region.

Keywords:

- Earthquake
- Attenuation
- BHRC
- Peak Ground Acceleration (PGA)
- Iran
- Logistic regression
- KNN- K Nearest Neighbours
- Random forest
- Ensemble model

1 Introduction

The evaluation of seismic hazard in a region of interest requires the use of ground-motion models. These models can be used to describe how a particular ground-motion parameter, such as spectral acceleration, attenuates with distance and varies for earthquakes of different sizes. Given a large number of strong-motion records, region-specific ground-motion models can be developed from empirical observations by regression analysis.

Here a model is developed to predict the PGA from the BHRC datasets containing all the information of the previous earthquakes

2 Literature Review

Iran is located in the Alpine-Himalayan seismic belt, as one of the most active tectonic regions of the world. Throughout history, the country has frequently suffered large and destructive earthquakes and experienced several major earthquakes in the past few decades. More than 70 percent of the big cities in Iran are located in the vicinity of seismic faults and in some cases, the active faults pass through the city.

3 Methodology: Research Methodology

Data Collection:

- Multiple **datasets are collected** from BHRC having various information about the series of earthquakes in Iran.
- The BHRC datasets were **merged** along with the *Earthquake Numbers.*

Data Cleaning:

- Rows with **missing time series** were removed from the merged dataset for better results.
- Columns with more than 40% missing data were dropped.
- Some statistical methods were used to fill the rest of the missing data.
- Some columns which are not necessary in predicting PGA were removed.

Model Training:

- The dataset was split into input part and the PGA value which was supposed to be predicted. The input part was standardised using *StandardScalar.*
- Now, the entire dataset was split into **test set** and **training set.**
- We trained the model using **Logistic Regression, K-Nearest Neighbours(KNN)** and **Random Forest Regression** using the training set.
- The trained models were tested with the test set to **predict the efficiency** of the model.
- Finally, two *Ensemble models* were created combining the previous models, one is the **Averaging** technique and the other is **Blending** technique.

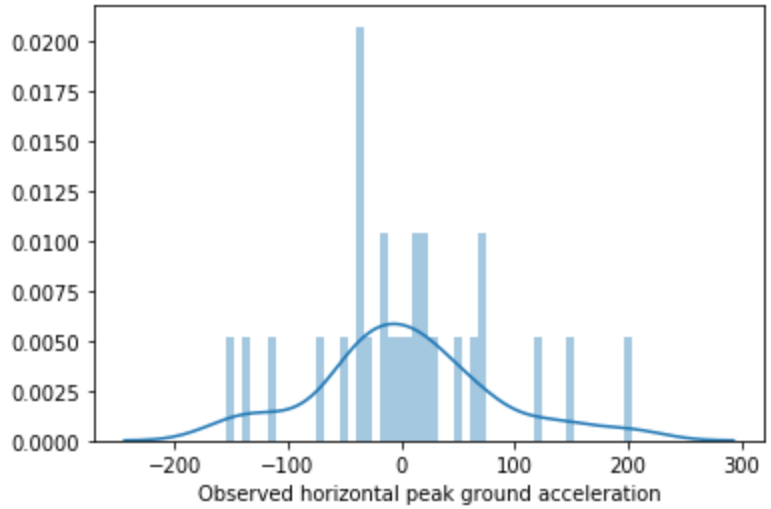
The efficiency of the models were predicted by calculating the Mean Average Error- ***MAE****, Mean Squared Error-* ***MSE*** *and Root Mean Squared Error- R****MSE. Lower the value, the higher is the efficiency.***

4 Analysis and Results

- **Logistic Regression Model :**

Prediction Efficiency

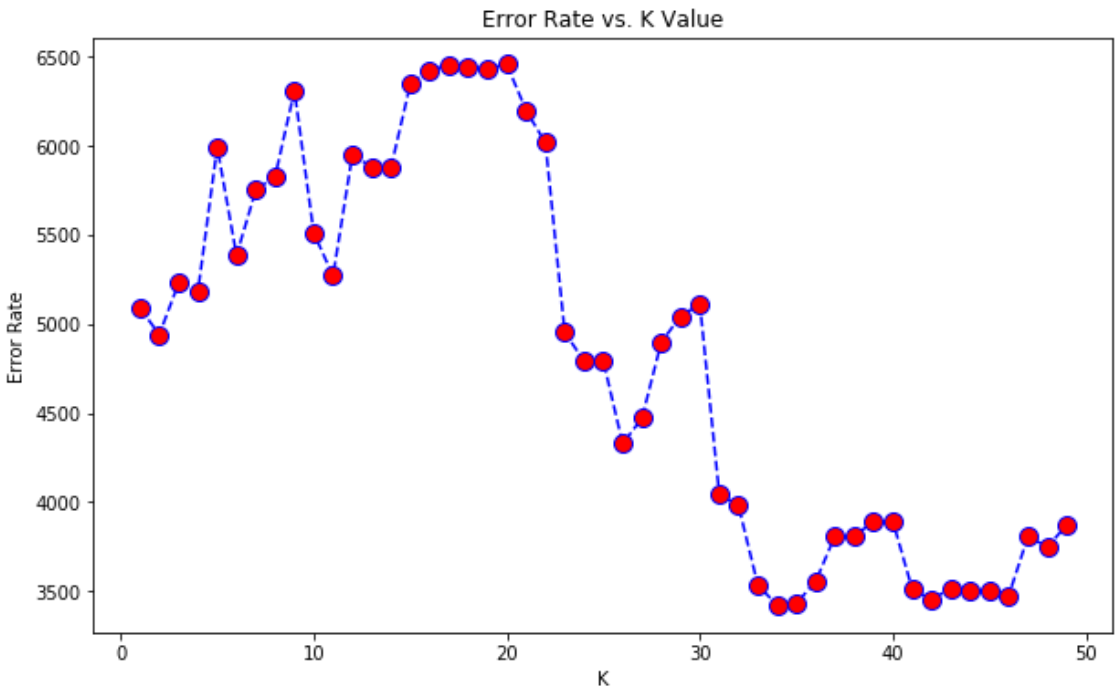
```
MAE: 58.666666666666664

MSE: 6148.740740740741

RMSE: 78.41390655196781
```



Here the difference between the predicted and the actual value is plotted. There is a peak near 0 difference.

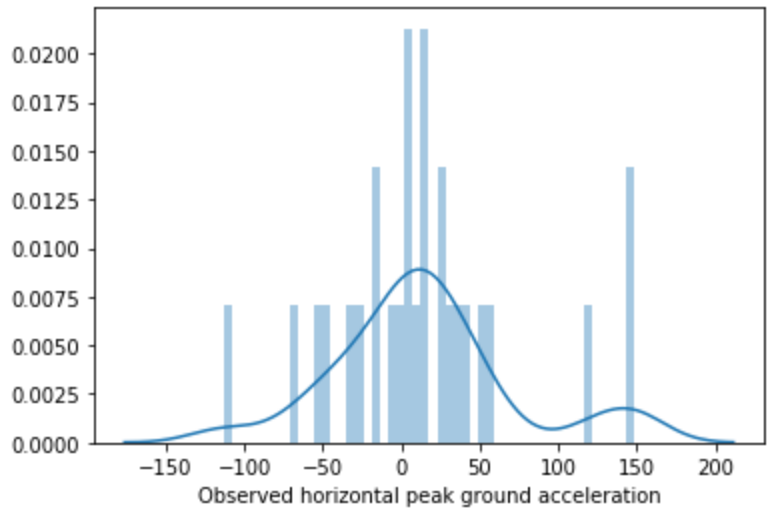
- **KNN Model :**

By plotting the error rate on the basis of RMSE on all values of K from 1 to 50. By looking at the plot below it seems that **K=33 is the most suitable one** for predicting the model.



Prediction Efficiency

```
MAE: 42.148148148148145

MSE: 3535.3333333333335

RMSE: 59.45866911841648
```



Here the difference between the predicted and the actual value is plotted. There is a peak near 0 difference.

- **Random Forest Classifier model :**

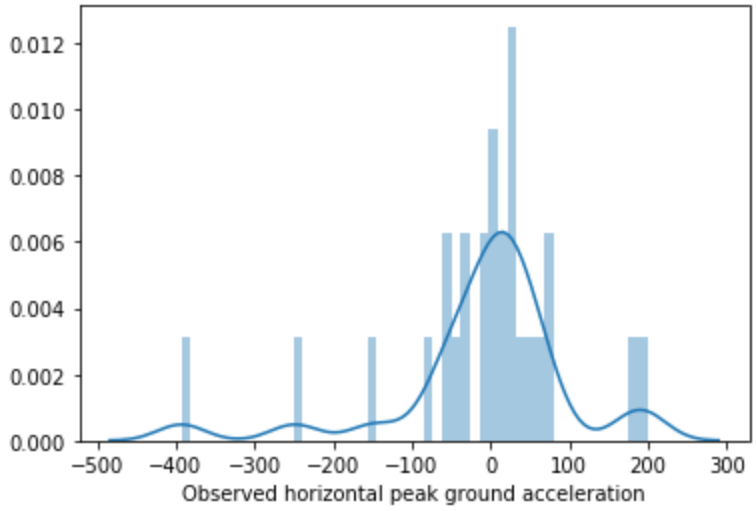
  100 estimators used in the classifier

  Prediction Efficiency :

  MAE: 70.66666666666667

  MSE: 12836.37037037037

  RMSE: 113.29770681867471



Here the difference between the predicted and the actual value is plotted. There is a peak near 0 difference.

- **Ensemble Model :**

Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model. A new model by combining all these three techniques above together (Ensemble Technique), so that we can get a more accurate model for our research work is built.

Here two ensemble models are built:

1. **Averaging Technique :** This model is built by calculating the mean of all the predicted values of those three models. Now these values are tested with the test set

   to calculate its efficiency.

   Prediction Efficiency:

MAE: 39.20987654320987

MSE: 3107.1687242798357

RMSE: 55.74198349789713

2. **Blending Method :** This model is built by splitting the entire dataset into **Test - Train - Validation sets** and model was built by training the model with logistic Regression on the results of the KNN and Random forest classifier.

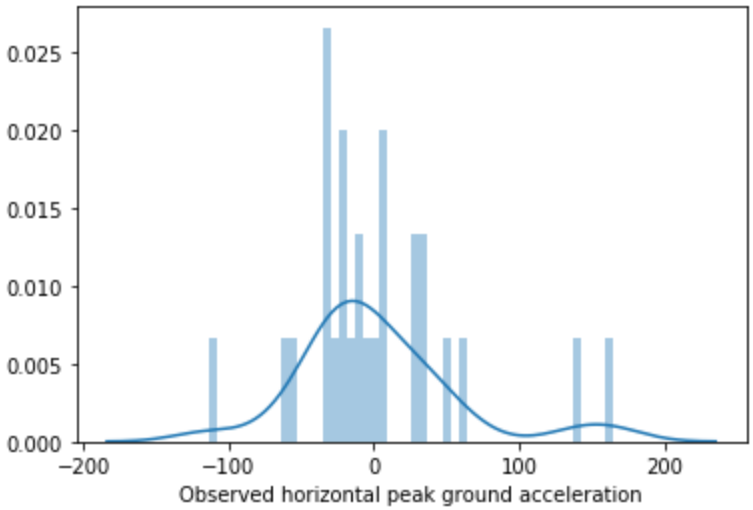
   Prediction Efficiency :

   MAE: 55.388888888888886

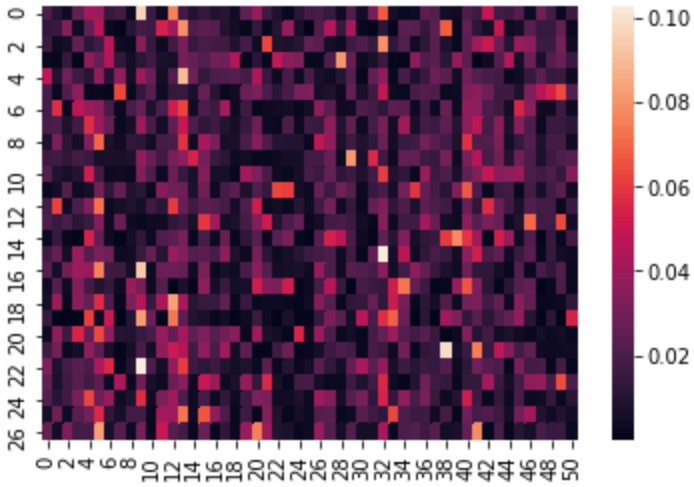
   MSE: 5441.055555555556

   RMSE: 73.76351100344638

Since the **Averaging Technique** is performing better than the **Blending Method,** so we are considering only **Averaging Technique.**



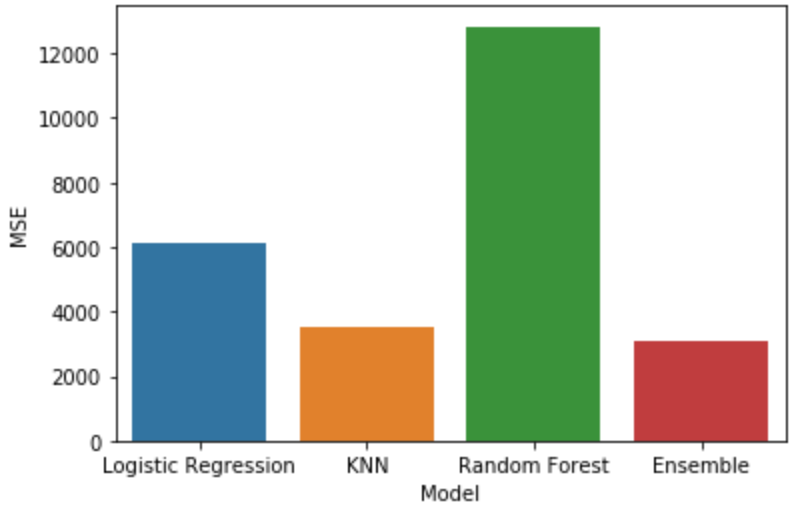
Here the difference between the predicted and the actual value is plotted. There is a peak near 0 difference.

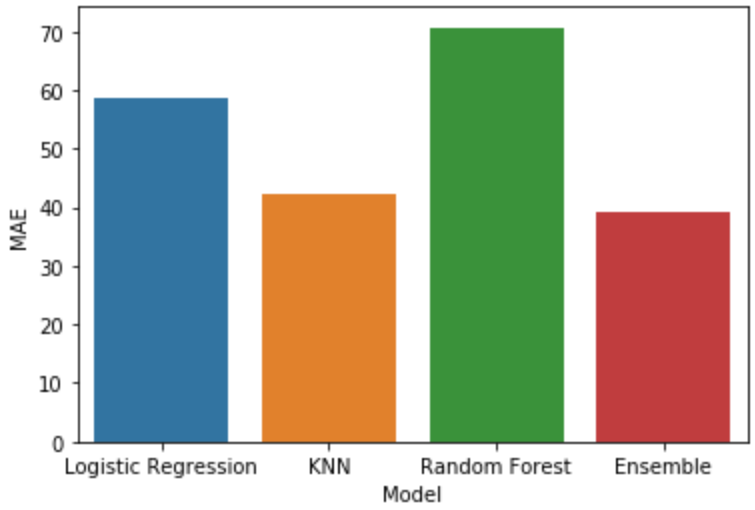


The Heat map is showing the probability for the occurrences in the test set in Ensemble method.

5 Conclusion

Here the Prediction Efficiencies of the different models are plotted. The lower is the error, the higher is its efficiency. So we can see that Ensemble model using the averaging technique is the best here. KNN model is also performing good, but Random Forest model is comparatively inefficient.





6 References

1. [An empirical spectral ground-motion model for Iran -H. Ghasemi · M. Zare ·Y. Fukushima · K.]
2. [An Evaluation of Disaster Preparedness in Four Major Earthquakes in Iran Y.O. Izadkhah 1* and K. Amini Hosseini]
3. *BHRC - Iran datasets attached.*
4. *Blackcoffer- Regional Attenuation-Final-Copy with TS.ipynb* ***attached***
5. *Final Analytics Ready Data.csv* ***attached***