Analysis of Earthquake Forecasting Using Supervised Machine Learning Classifiers

Presented By
Chinmayi Rane (2019220072)

OBJECTIVE

- To implement various supervised machine learning algorithms on an earthquake dataset

- To determine the correctness of model information based on accuracy, precision, recall, Mathew's coefficient and F1 score

- To classify the earthquake instances from the dataset based on fatal, mild and moderate earthquakes

Literature Survey & Research gaps

Reference: *Rikitake, T. Classification of earthquake precursors. Tectonophysics 1979, 54, 293–309*

Precursors	Theory
Anomalous animal behaviour	Abnormal behavior and more intensive responses of animals are observed during the high magnitude of the earthquake (5 or more). These responses are mostly observed in the epicentral region—close to the active faults
Hydrochemical Precursors	During the seismically inactive period, it has been observed that absorption levels of deliquescing minerals and gassy integrands of underground water in a seismically active region remain almost constant
Temperature change	A tolerable rise of temperature by 10 ◦C and 15 ◦C occurs before earthquakes

Literature Survey & Research gaps

Paper Title	**Year of publication**	**Authors**	**Description**
Neural Network Applications in Earthquake Prediction (1994–2019): Meta-Analytic and Statistical Insights on Their Limitations	2020	Arnaud Mignan and Marco Broccardo	Used Artificial Neural Networks. The black-box nature of an ANN and its high variance can easily lead to fallacious physical interpretations. Potential precursors were not taken into consideration.
Earthquake Early Warning: Recent Advances and Perspectives	2020	Gemma Cremen and Carmine Galassoy	Have discussed various approaches to detection of the earthquake and estimating it magnitude

Literature Survey & Research gaps

Paper Title	Year of publication	Authors	Description
Geo-sensor(s) for potential prediction of earthquakes: can earthquake be predicted by abnormal animal phenomena?	2018	Kai Cao & Qunying Huang	Used ‘citizen-as-sensors’ to locate anomalous animal behaviours, which prove to be reliable signals for a potential earthquake
A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility	2016	Wei Chen, Xiaoshen Xie, JialeWang, Biswajeet Pradhan, Haoyuan Hong, Dieu Tien Bui, Zhao Duan, Jianquan Ma	3 models were compared – Random forest, logistic model tree and classification and regression tree

Proposed Methodology

Dataset and Tools

- The original dataset has 21 attributes and 1 target variable.
- Few of the attributes are -
  - Time, Latitude, Longitude, Depth, number of seismic stations etc
- Target variable (magnitude) was divided into 3 classes -
  - Fatal Earthquakes (Magnitude > 5.5)
  - Moderate Earthquakes (5.5 > Magnitude > 4.5)
  - Mild Earthquakes (Magnitude < 4.5)

Dataset taken from: https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php

Dataset and Tools

No	Attribute Name	Description
1	time	Time when the earthquake occurred in the yyyy-mm-dd HH:MM: SS format
2	latitude	Latitude of the place
3	longitude	Longitude of the place
4	depth	Depth of the earthquake in kilometres
5	nst	The number of Seismic stations, which is used to determine the earthquake location
6	gap	Seismic Gap in degree (0 to 180 degrees)
7	dmin	Horizontal distance between epicentre and nearest station in degrees
8	rms	The root mean square of the travel time residual.
9	net	Data contributor ID
10	id	Database id of record
11	updated	Most recently updated time of earthquake
12	place	Description of Geographical position
13	type	Type of Seismic Event ("Earthquake", "Quarry")
14	locationSource	Name of the network that reported the location of the earthquake
15	magSource	Name of the network that reported the magnitude of the earthquake
16	horizontalError	Horizontal error of the location in kilometres
17	depthError	Depth error of the location in kilometres
18	magError	Standard error of the magnitude
19	magNst	The earthquake magnitude, which is determined using the number of Seismic stations
20	status	Indicates that the earthquake was reviewed by humans
21	Class	Target Variable (Fatal Earthquake, Moderate Earthquake and Mild Earthquake)Fatal Earthquake Magnitude Value > 5.5, Moderate Earthquake Magnitude Value between 4.5 and 5.5,Mild Earthquake Magnitude Value between 2.5 and 4.5

Preprocessing/Data Visualization

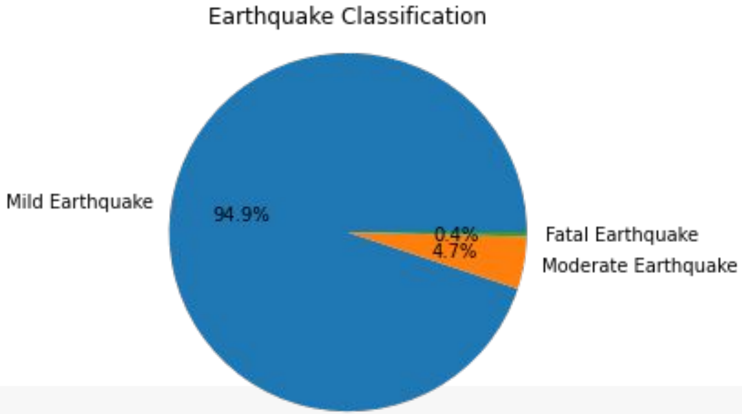
```
df['class'].value_counts()
```

```
1.0     8415
2.0      419
3.0       33
Name: class, dtype: int64
```

8415 Mild earthquakes

419 Moderate earthquakes

33 Severe earthquakes are present in the dataset.

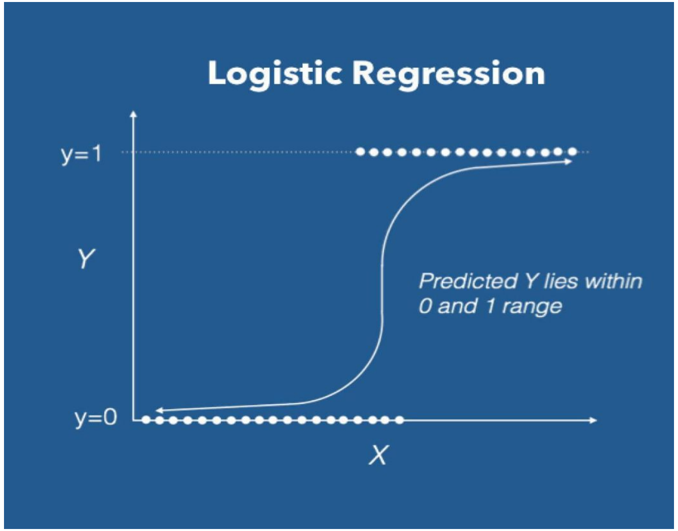



```
df.head()
```

	latitude	longitude	depth	mag	nst	gap	dmin	rms	horizontalError	depthError	magError	magNst	class
0	33.190167	-115.595833	1.78	1.72	21.0	81.0	0.02089	0.19	0.31	0.42	0.129	27.0	1.0
1	61.662800	-150.393800	18.00	1.70	NaN	NaN	NaN	0.34	NaN	0.40	NaN	NaN	1.0
2	33.197333	-115.602500	1.42	1.47	19.0	88.0	0.02200	0.25	0.42	0.59	0.193	27.0	1.0
3	34.067500	-117.259000	10.68	0.95	27.0	71.0	0.12170	0.23	0.35	1.04	0.126	26.0	1.0
4	36.530167	-117.971664	6.81	3.00	45.0	55.0	0.10040	0.24	0.30	0.85	0.268	24.0	1.0

ML Algorithms

1. ZeroR
   - ZeroR depends on the relevant target and disregards all other relevant predictors.
   - It creates a frequency table of the target variable and selects the most frequent value.
2. Logistic Regression
   - It is a predictive algorithm and works based on the concept of a probability model.
   - Logistic regression works using a logistic function

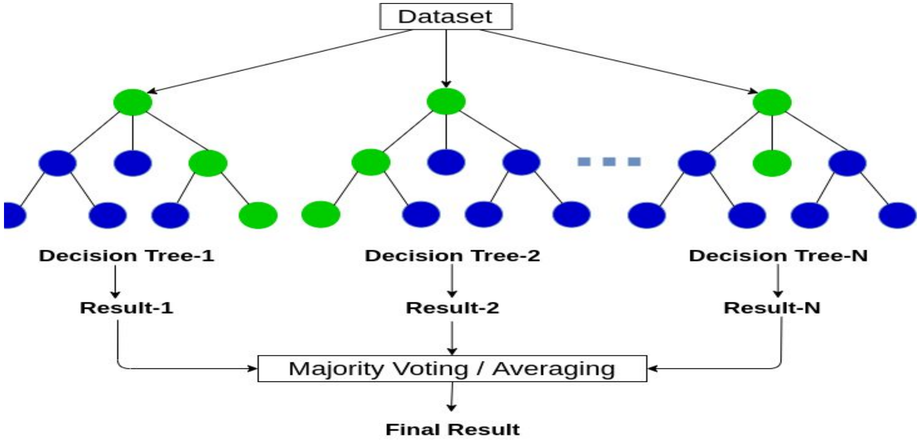


ML Algorithms

3. Random Forest
    - Type of supervised machine algorithm
    - It creates a random decision tree and merges it into a single forest.
    - Selects a random feature for improving the performance of its accuracy
4. Linear Model Tree
    - Linear Trees combine the learning ability of Decision Tree with the predictive and explicative power of Linear Models.
    - The data are split according to simple decision rules. The goodness of slits is evaluated in gain terms fitting Linear Models in the nodes.

ML Algorithms

5. Decision Tree
    - Designed by using a stochastic process
    - Each node of the tree represents an input attribute. Number of edges of the tree is equal to the possible number of input attributes.



Result Analysis

Result Analysis

	Accuracy Score	Precision Score	Recall Score	F1 score	Matthews Coefficient
Zero R	0.94757	0.31585	0.33333	0.32436	0.0
Linear Tree	0.98139	0.59905	0.59781	0.59842	0.80476
Logistic Regression	0.9915	0.95891	0.76091	0.80386	0.91574
Decision Tree	1.0	1.0	1.0	1.0	1.0
Random Forest	1.0	1.0	1.0	1.0	1.0

Conclusion

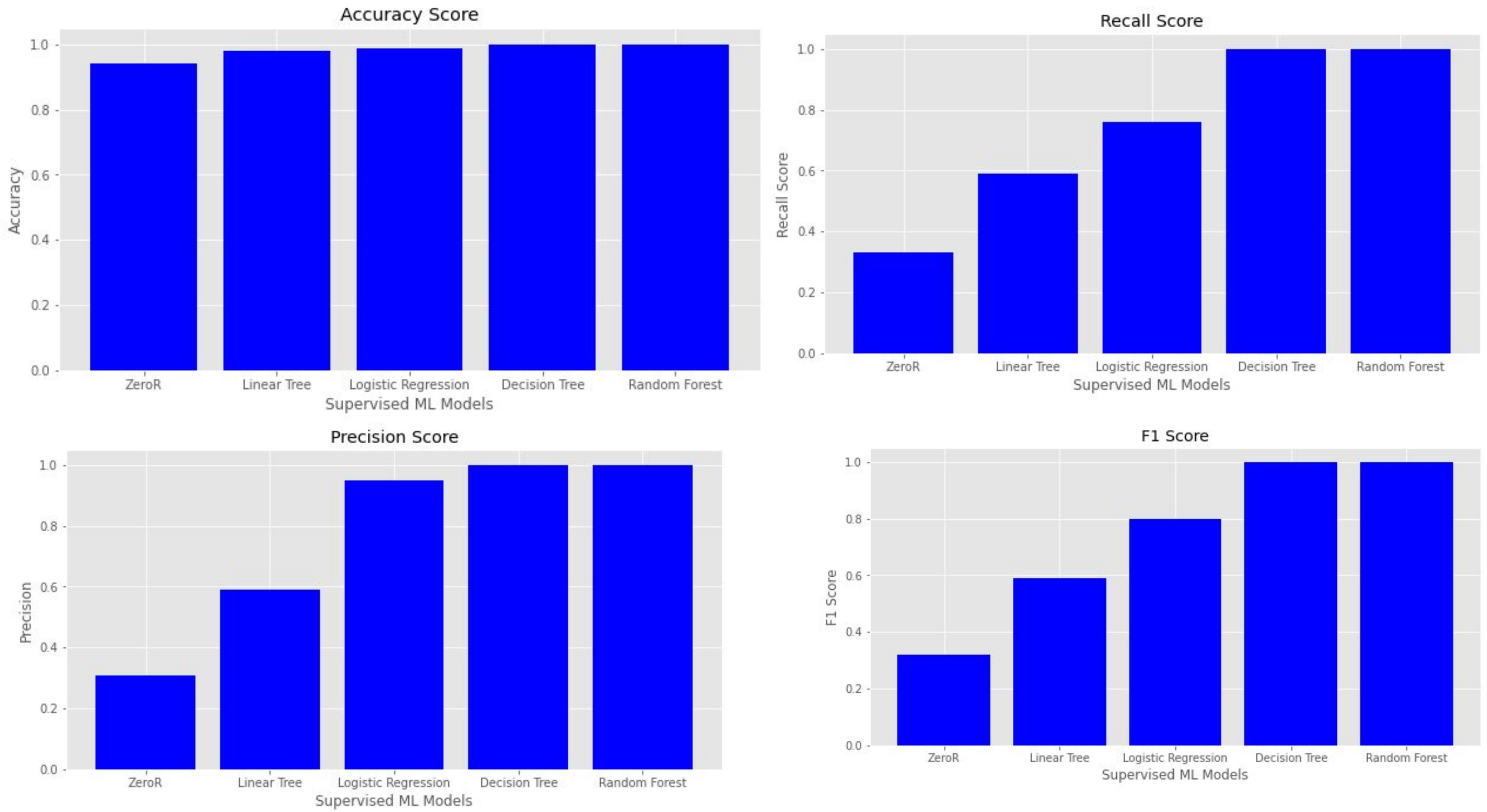
The research aim was to find earthquake magnitude ranges. From the results, it was observed that any future earthquake's magnitude range can be predicted.

Seven machine learning classifiers were applied to an earthquakes dataset. The comparisons between them were noted.

As earthquakes are shifts in the tectonic plates, the magnitude in which earthquakes occur would not vary much in a region

The results do not say anything about being able to predict when an earthquake will occur. Only the magnitude ranges of future earthquakes that will occur someday.

From the project, we can conclude that Random Forest has performed very well on the given dataset


Accuracy Score
Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
ZeroR
Linear Tree
Logistic Regression
Decision Tree
Random Forest
Supervised ML Models
Recall Score
Recall Score
Precision Score
Precision
F1 Score
F1 Score

Thank You