SIGNIFICANT OF EARTHQUAKE HAPPENS USING DATASCIENCE

[1]<*designation of author* >

[#]<*designation of author*>

<example@example.com>

ABSTRACT :

An earthquake is the shaking of the surface of the Earth resulting from a sudden release of energy in the Earth's lithosphere that creates seismic waves. At the Earth's surface, earthquakes manifest themselves by shaking and displacing or disrupting the ground. So predicting the factors of an earthquake is a challenging job as an earthquake does not show specific patterns resulting in inaccurate predictions. Techniques based on machine learning are well known for their capability to find hidden patterns in data. The machine learning model is built based on the past data related to earthquakes where the model can learn the pattern from the data and takes the consideration of the factors. The factors which are taken into consideration are the pre-processed data that is the dependency of the factors are checked in accordance with earthquake. Comparison of machine learningalgorithms are done for better prediction and performance metrics are also calculated an evaluated.

INTRODUCTION :

While mining this data set through normal EDA process I came across the fact that not all earthquakes are natural and few are indeed caused by humans although very small numbers.At the end of the analysis I have tried to predict earthquakes and other quakes (seismicactivities related to explosion, quarry blast etc.). I have also tried to handle the class imbalance problem because the data set is 98:2.The next steps are pretty usual ones with loading and probing the data. Let's get started with loading the libraries first.

LITERATURE SURVEY:

In [1] PTWP and MTR tasks for the EEWS to determine earthquake source parameters in real time. The N-check algorithm improves window prediction selection to reduce false alarms in multistation waveforms that have noise, and MTR with hard-shared orthogonal are proven to improve earthquake parameter determination performance. The response time that is the prediction need to be improved. Classification and confusion matrix are not calculated. Inpaper[2], Supporting earthquake risk management with clear seismic communication may necessitate encounters with various popular misapprehensions regarding earthquake prediction. Drawing on technical data as well as insights from anthropology and economics, this paper addresses common and scientifically-unsupported ideas about earthquake prediction, as well as the state of science-based studies regarding statistical forecasting and physical precursors.Low computation speed.In [3] The earthquake sequence has revealed unique issues and complexities for the owners of commercial and multi-storey residential buildings in relation to unexpected technical, legal, and financial challenges when making decisions regarding the future of their buildings impacted by the earthquakes. The paper presents a framework to understand the factors influencing post-earthquake decisions (repair or demolish) on multi-storey concrete buildings in Christchurch. The study, conducted in 2014, includes in-depth investigations on 15 case-study buildings

using 27 semi-structured interviews with various property owners, property managers, insurers, engineers, and government authorities in New Zealand. It does not handle a complex problem.

LOGISTIC REGRESSION :

It is a statistical method for analysing a data set in which there are one or more independent variables that determine an outcome. The outcome is measured with a di- chotomous variable (in which there are only two possible outcomes). The goal of logistic regression is to find the best fitting model to describe the relationship between the dichot-omous characteristic of interest and a set of independent variables. Logistic regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1.

RANDOM FOREST:

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of de- cision trees at training time and outputting the class that is the mode of the classes (clas- sification) or mean prediction (regression) of the individual trees. Random decision for- ests correct for decision trees' habit of over fitting to their training set. Random forest is a type of supervised machine learning algorithm based on ensemble learning. Ensemble learning is a type of learning where you join different types of algorithms or same algo- rithm multiple times to form a more powerful prediction model.

NAÏVE BAYES :

The Naive Bayes algorithm is an intuitive method that uses the probabilities of each attribute belonging to each class to make a prediction. It is the supervised learning approach you would come up with if you wanted to model a predictive modeling problem probabilistically. Naive bayes simplifies the calculation of probabilities by assuming that the prob- ability of each attribute belonging to a given class value is independent of all other attributes. This is a strong assumption but results in a fast and effective method.

DECISION TREE CLASSIFIER:

It is one of the most powerful and popular algorithm. Decision-tree algorithm falls under the category of supervised learning algorithms At the beginning, we consider the whole training set as the root.Attributes are assumed to be categorical for information gain, attributes are as- sumed to be continuous.On the basis of attribute values records are distributed recursively.It utilizes an if-then rule set which is mutually exclusive and exhaustive for classification.

EXISTING SYSTEM :

They proposed two methods and named them as picking target window prediction PTWP and multitarget regression (MTR) tasks to determine earthquake source parameters. The N-check algorithm improves window prediction selection to reduce false alarms in multistation waveforms that have noise, and MTR with hard-shared orthogonal are proven to improve earthquake parameter determination performance. Our system can provide reliable earthquake parameters. The three stations with three- component seismogram traces are represented in red, green, and blue components to form pixels in a row in one frame in a 10 s window. The sampling rate at each station varied between 20 and 25 Hz and was then normalized to 20 Hz. They used a band pass filter to minimize noise and normalize each stream by dividing its absolute peak amplitude. The data set has high noise for 506 seismic events and has a peak SNR of less than 50 dB.

PROPOSED SYSTEM :

Earthquakes are a natural disaster that can cause a lot of

damage to both lives and properties. The machine learning is applied to every field where the dataset can be used to learn patterns and then from that pattern the prediction can be done. Our objective is to build a machine learning model that uses the past earthquake related dataset the data is pre-processed by using variable identification that is finding the dependent and independent variables after that the data is used to train the during machine learning libraries. Different algorithms are used to compare the model and the performance metrics are calculated and evaluat

RESULT :

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

In this project The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on public test set is higher accuracy score is will be find out. This application can help to find the Prediction of Earth Quake.

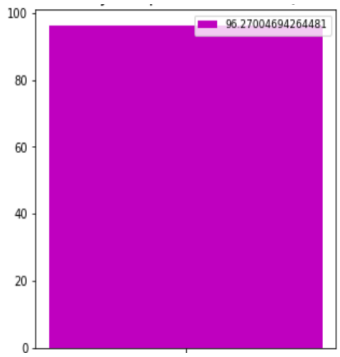

Fig 1:High Accuracy for Earthquake prediction

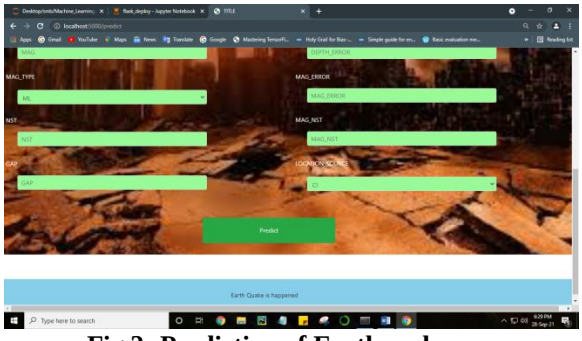

Fig 2: Prediction of Earthquake

CONCLUSION:

Therefore, with further optimization of the earthquake dataset and improvements in the algorithm,a computer-aided detection system can be expected to become an effective and efficient method of significant of Earthquake happens using Datascience concept the second module Random Forest gave the more accuracy compared to other modules.So by using Random Forest we train the machine for more accuracy.Hence we demonstrated that the DataScience and Machine learning Algorithm was useful for assessing the diagnosis and predictability.

In Future Work, Earth Quake prediction to connect with AI model.To automate this process by show the prediction result in web application. To optimize the work to implement in Artificial Intelligence environment

REFERENCE :

[1] C. Pratama, David P. Sahara, L. S. Heliani, S. Rasyid,Zharfan Akbar, Faiz Muttaqy, and Ajat Sudrajat(2021)," Earthquake Early Warning System Using Ncheck and Hard-Shared Orthogonal Multitarget Regression on Deep Learning"doi: https://doi.org/10.1109/LGRS.2021.3066346.

[2]S. Widiyantoro et al., (2020)"Implications for megathrust earthquakes and tsunamis from seismic gaps south of Java Indonesia," Sci. Rep., vol. 10,no. 1, pp. 1-11Dec.2020,doi:10.1038/s41598-020-72142-

z.

[3]X. Zhang, J. Zhang, C. Yuan, S. Liu, Z. Chen, and W. Li,(2020) “Locating induced earthquakes with a network of seismic stations in Oklahoma via a deep learning method,” Sci. Rep., vol. 10, no. 1, pp. 1—12, Dec. 2020,doi: 10.1038/s41598-020-58908-5.

[4]D. Jozinovi´c, A. Lomax, I. Štajduhar, and A. Michelini, “Rapid predic-tion of earthquake ground shaking intensity using raw waveform data and a convolutional neural network,” 2020, arXiv:2002.06893. [Online].Available: https://arxiv.org/abs/2002.06893

[5]M. P. A. van den Ende and J. P. Ampuero, “Automated seismic source characterization using deep graph neural networks,” Geophys. Res. Lett., vol. 47, no. 17, pp. 1—11, Sep. 2020, doi: 10.1029/2020GL088690

[6]O. M. Saad and Y. Chen,(2020) “Earthquake detection and P-wave arrival time picking using capsule neural network,” IEEE Trans. Geosci.Remote Sens., early access, Sep. 28, 2020, doi: 10.1109/tgrs.2020.3019520.

[7] S. M. Mousavi, W. L. Ellsworth, W. Zhu, L. Y. Chuang, andG. C Beroza,(2020)“Earthquake transformer—An attentive deep-learningmodel for simultaneous earthquake detection and phase picking,” NatureCommun., vol. 11, no. 1, pp. 1—12, Dec. 2020, doi: 10.1038/s41467-020-17591-w.

[8] Z. E. Ross, Y. Yue, M. Meier, E. Hauksson, and T. H. Heaton,(2019)“PhaseLink: A deep learning approach to seismic phase association,”J. Geophys. Res., Solid Earth, vol. 124, no. 1, pp.856—869,Jan.2019,doi:10.1029/2018JB016674.

[9]E. Pardo, C. Garfias, and N. Malpica,(2019) “Seismic phase picking using convolutional networks,” IEEE Trans. Geosci. Remote Sens., vol. 57,no. 9, pp. 7086—7092, Sep. 2019, doi: 10.1109/TGRS.2019.2911402

[10]A. Lomax, A. Michelini, and D. Jozinovi´c,(2019) “An investigation of rapid earthquake characterization using single-station waveforms and a con-volutional neural network,” Seismolog. Res. Lett., vol. 90, no. 2A,pp. 517—529, Mar. 2019, doi: 10.1785/0220180311

[11]M. Kriegerowski, G. M. Petersen, H. Vasyura-Bathke, and M. Ohrnberger, (2019)“A deep convolutional neural network for localization of clustered earthquakes based on multistation full waveforms,” Seismolog. Res. Lett., vol. 90, no. 2A, pp. 510—516, Mar. 2019, doi: 10.1785/0220180320

[12] A. Howard et al.,(2019) “Searching for mobileNetV3,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 1314—1324, doi:10.1109/ICCV.2019.00140.

[13]M. Tan and Q. V. Le,(2019) “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. 36th Int. Conf. Mach. Learn (ICML), Jun. 2019, pp. 10691—10700.