PROJECT SPARTA DATA SCIENTIST CAPSTONE PROJECT

IN

PHILIPPINE EARTHQUAKE PREDICTION MACHINE LEARNING MODEL

November 16, 2024

Edmar C. Beatingo

Introduction

According to the United States Geological Survey (2015), an earthquake is a sudden movement of the Earth, caused by the abrupt release of strain that has accumulated over a long time. Earthquake is the passage of vibrations (seismic wave) that spread out in all directions from the source of the disturbance when rocks are suddenly disturbed (Adagunodo 2015 and Sunmonu 2015). Moreover, for the Philippine Institute of Volcanology and Seismology (2018), earthquake is a weak to violent shaking of the ground produced by the sudden movement of rock materials below the earth's surface. The Philippines is located in one of the most seismically active regions of Asia (Rusydy et al, 2018). It is therefore beneficial to develop pattern analysis and earthquake models to predict earthquakes (Mejia et al, 2015).

Problem Statement and Objectives

The Philippines is highly susceptible to both geophysical and climate-related disasters (Bollettino, et al, 2020). Also, as the islands are located within the "Ring of Fire" between the Eurasian and Pacific tectonic plates, earthquakes and volcanoes are posing serious risks to the safety of the populace (UNDDR 2019). With continued development and population growth in hazard-prone areas, it is expected that damage to infrastructure and human losses would persist and even rise unless appropriate measures are immediately implemented by the government (Lagmay et al, 2017).

This study aims to analyze earthquake data provided by the Philippine Institute of Volcanology and Seismology (PHIVOLCS) using machine learning algorithms to predict the recurrence of earthquakes in specific locations. The dataset available on Kaggle, titled "Philippine Earthquakes from PHIVOLCS," offers a structured collection of seismic events recorded by PHIVOLCS from 2016 onwards. It includes columns such as date, time, magnitude, depth, location, and coordinates (latitude and longitude), similar to the information presented on the PHIVOLCS website. The predictive model aims to identify patterns, trends, and high-risk areas, thereby assisting in disaster preparedness and response strategies.

Analysis

The entire data manipulation step of the analysis utilized the Python programming language. Python is a powerful high-level, interpreted, interactive, and object-oriented scripting language created by Guido Van Rossum in late 1980’s (Saabith et al, 2019). Python is a popular choice for data science because it is easy to use for data manipulation and analysis. (Abu and Gross, 2023). All codes for this study were saved in a text file format using the Jupyter Notebook file extension for Python: earthquake_prediction_notebook.ipynb.

The raw dataset phivolcs_earthquake_data_raw.csv was obtained from Kaggle.com, credited to Bwandowando, who extracted Philippine earthquake data from PHIVOLCS starting in 2016. The dataset was then cleaned through various steps. Data cleaning is a process used to determine inaccurate, incomplete or unreasonable data and then improve the quality through correcting of detected errors and omissions (Singh, 2023).
Data cleaning steps done:

1. Check descriptive statistics and data types of all columns.
2. Dropped rows with any missing values.
3. Convert date column to date format.
4. Convert Latitude, Longitude, Depth_In_Km, and Magnitude columns to numeric.
5. Convert Latitude, Longitude and Magnitude columns to float.
6. Convert Depth_In_Km column to integer.
7. Round Latitude, Longitude, and Magnitude columns to 2 decimal places.
8. Drop rows with any NaN values that resulted from the conversion and removed duplicates.

The cleaned raw dataset was then saved to a new CSV file named phivolcs_earthquake_data_cleaned.csv.

The cleaned dataset was pre-processed using feature engineering for exploration. Feature engineering is the process of transforming raw data into relevant information for use by machine learning models (Murel and Kavlakoglu, 2024).
Feature engineering steps done:

1. Extract new feature columns (Year, Month, Day and Hour) from Date_Time_PH columns.
2. Split the 'Location' column into 'Distance', 'Direction', and 'Place'
3. Save the updated dataframe to a new CSV file named earthquake_new_features.csv.
4. Convert month and day columns to string equivalent and create new columns 'Month_Str' and 'Day_Str'.

5. Save the updated dataframe to a new CSV file named earthquake_added_features.csv.
6. Normalized place names by removing spaces and converting to lowercase.
7. Extract town and province from normalized place column.
8. Identify anomalies by checking and dropping duplicates.
9. Save the updated dataframe to a new CSV file named earthquake_updated_features.csv.

The updated feature dataset is now ready for exploration. Exploratory data analysis (EDA) is an approach using descriptive statistics and graphical tools to better understand data (Camizuli and Carranza, 2018).

Exploratory data analysis steps done:

1. Drop any rows with missing values.
2. Convert 'Year', 'Month', 'Day', and 'Hour' columns to integer type.
3. Convert 'Distance' to float type.
4. Convert 'Direction' and 'Place' to string type.
5. Clean the 'Province' column by removing leading/trailing spaces, extra parenthesis characters, and converting to proper case.
6. Clean the 'Town' column by removing leading/trailing spaces, extra parenthesis characters, and converting to proper case.
7. Save the cleaned dataframe to a new CSV file named earthquake_cleaned_features.csv.

The cleaned feature dataset is then being set for anomaly detection testing the town and province column. The Anomaly Detection, Classification and Identification Tool (ADCIT) is an open source Matlab and Python code used for detection, classification and identification of anomalies in power system state estimation (Asefi, 2023).

Anomaly detection testings done:

1. Define a list of known valid provinces and towns.
2. Correct detected anomalies of identified provinces and towns.
3. Save the dataframe to a new CSV file named earthquake_preprocessed_features.csv.
4. Save the dataframe to a new CSV file named town_anomalies.csv

The preprocessed features dataset is then visualized. Python is a good choice, providing a wealth of third-party libraries, open-source communities, and continuously optimized documentation for data visualization (Cao et al, 2021).

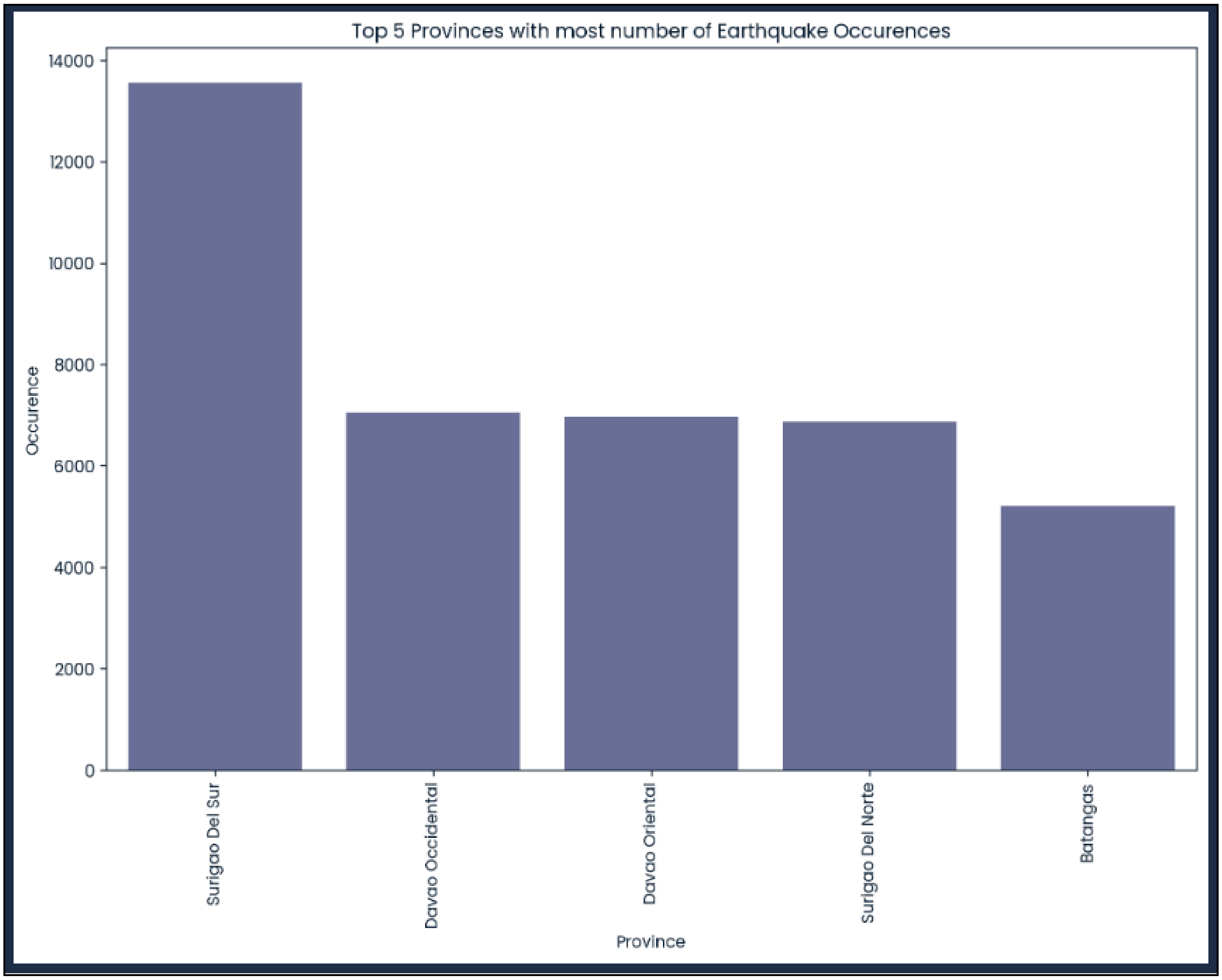

Chart No.1 The chart shows the provinces with the highest number of earthquake occurrences, highlighting the significant seismic activity in these areas.

Insights:

Surigao Del Sur has the highest number of recorded earthquakes in the Philippines, largely due to its location on the Philippine Fault Zone. Notably, four of the top five earthquake-prone provinces (Surigao Del Sur, Davao Occidental, Davao Oriental, and Surigao Del Norte) are in Mindanao, emphasizing the region's tectonic activity. Batangas is the only province from Luzon among the top five, primarily affected by the West Valley Fault and Taal Volcano.

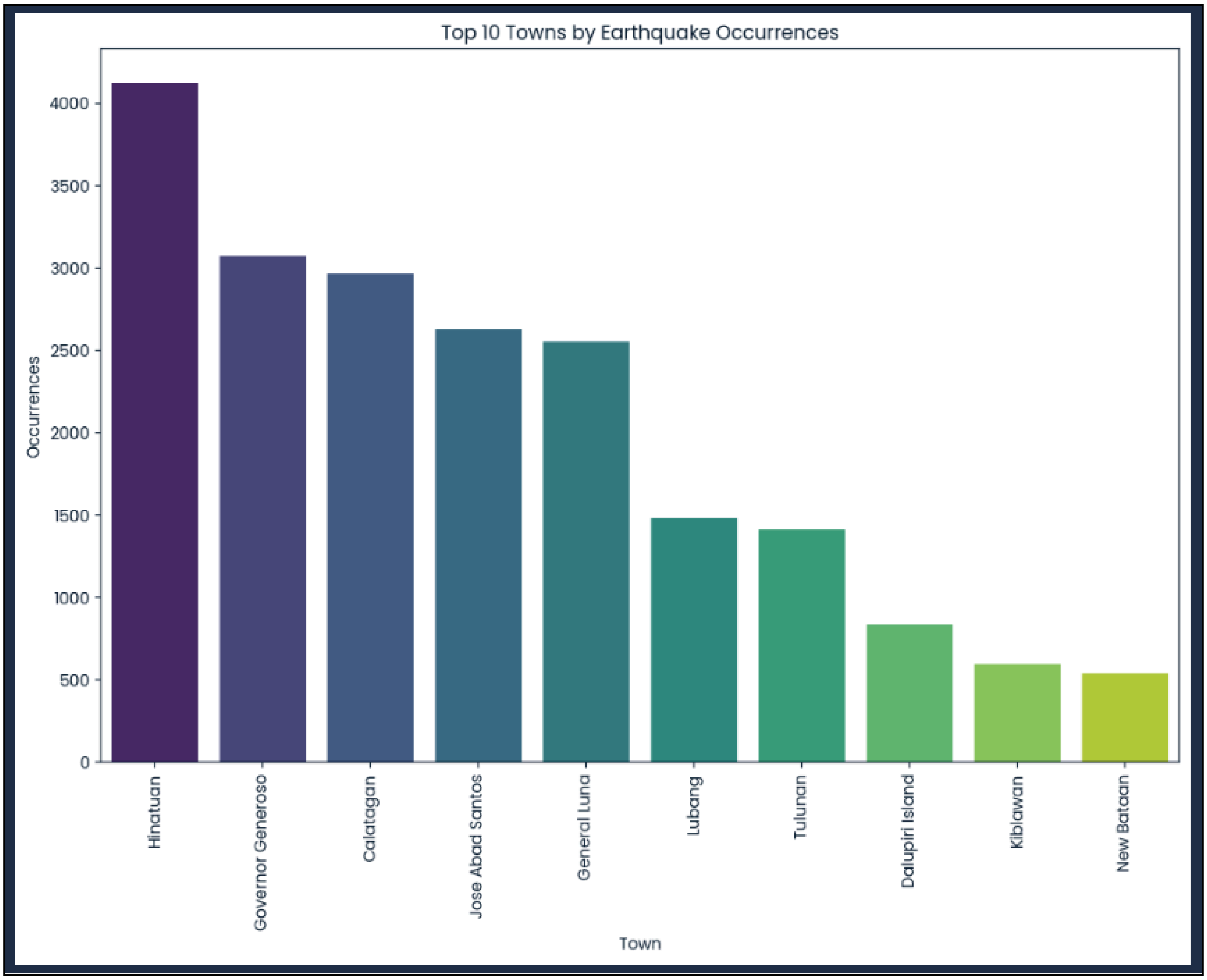

Chart No.2 The chart shows the top 10 towns with the highest number of earthquake occurrences.

Insights:

Hinatuan has the highest number of earthquake occurrences in the region, exceeding 4,000, likely due to its location near active seismic zones like the Philippine Trench. Other towns such as Governor Generoso, General Luna, and Jose Abad Santos also experience significant seismic activity, indicating a clustering of earthquake-prone areas in southeastern Mindanao. In contrast, towns like Tulunan, known for recent high-magnitude earthquakes, highlight the need for preparedness due to their vulnerability. Meanwhile, Kiblawan and New Bataan show lower earthquake counts, which may suggest either less tectonic activity or underreporting.

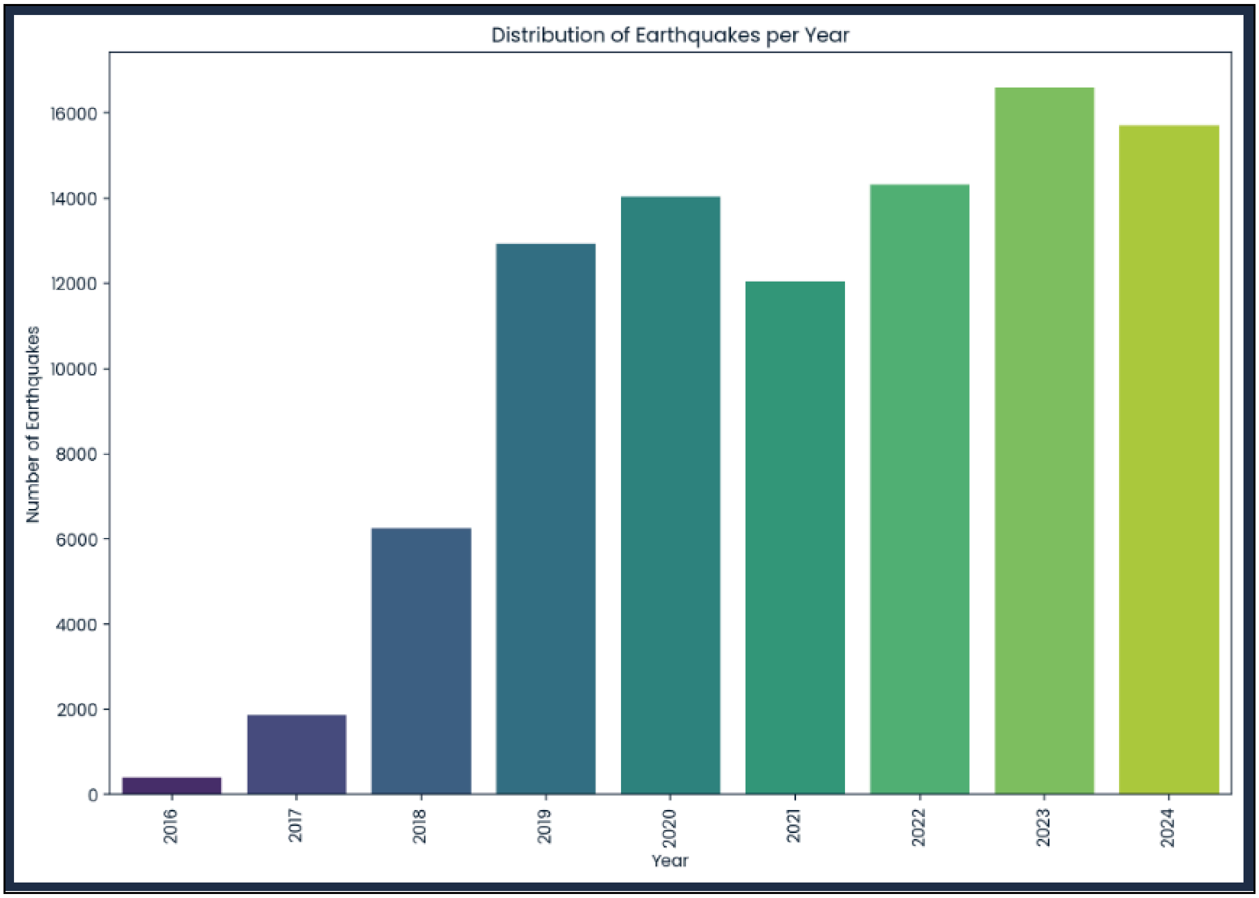

Chart No.3 Shows the distribution and trends displayed in the chart showing earthquakes by year.

Insights:
There has been a notable increase in recorded earthquakes from 2016 to 2024, with a significant rise starting in 2019. Earthquakes have stabilized at over 12,000 occurrences annually since then, peaking in 2023 before a slight decline in 2024. This increase may reflect enhanced monitoring systems rather than a true rise in activity.

The preprocessed dataset of the season features are being normalized to further now how it affects one another. Normalization is the process of casting the data to the specific range, like between 0 and 1 or between -1 and +1 (Ali et al, 2014). Rainy season was defined from the month of June to November and dry season from December to May. The steps and changes made to the dataset are then updated original dataset with the aggregated features named earthquake_aggregated_features.csv.

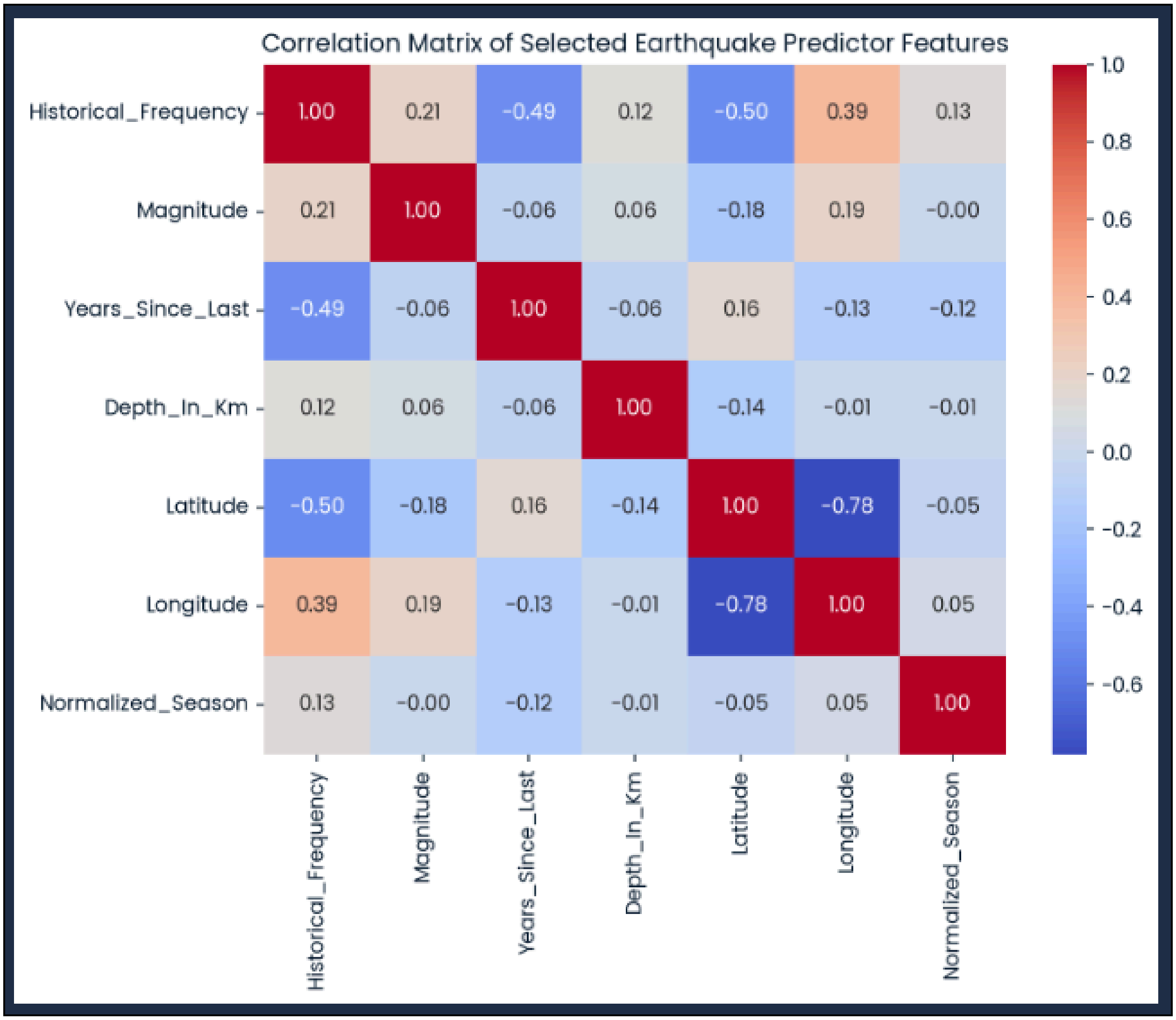

Chart No.4 Shows the correlation of defined features related to earthquakes from the aggregated features.

Correlation identifies the association between variables (Boussiala 2021).

Insights:

The graph implies that high historical frequency of earthquakes is strongly associated with larger magnitudes. There's a slight positive correlation between historical frequency and normalized seasons, suggesting some seasonal influence on seismic activity. Additionally, areas with frequent historical earthquakes tend to have shorter intervals between events. However, factors like the magnitude of earthquakes, years since the last quake, and depth show weak correlations with other features, indicating they may not be strong predictors of earthquake characteristics. Overall, historical frequency emerges as a key predictor for earthquake magnitude.

The combination of machine learning libraries and flexibility makes Python uniquely well-suited to developing sophisticated models and prediction engines that plug directly into production systems (Kostyuchenko and Gosudarev, 2018).
Data modeling steps done:

1. Created a target variable for earthquakes with magnitude ≥ 5.5.
2. Define features and target with magnitude column.
3. One-hot encode categorical features.
4. Split the data into training and testing sets.
5. Initialize and train the Random Forest Classifier.
6. Predict probabilities for the test set and evaluate model.
7. Display the ROC AUC score and classification report.

This study used the Random Forest model. The Random Forest is appropriate for high dimensional data modeling because it can handle missing values and can handle continuous, categorical and binary data (Ali et al, 2012).

Refer to earthquake_prediction_notebook.ipynb for the codes in Python.

```
ROC AUC Score: 0.7066
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	18800
1	0.00	0.00	0.00	34
accuracy			1.00	18834
macro avg	0.50	0.50	0.50	18834
weighted avg	1.00	1.00	1.00	18834

Table No. 1 Shows the ROC AUC score for model evaluation which is a measure of the model's ability to distinguish between classes.

Insights:

A score of 0.5 indicates no discrimination (i.e., random guessing), while a score of 1.0 indicates perfect discrimination. In this case, a score of 0.7066 suggests that the model has a good ability to distinguish between earthquakes with magnitude ≥ 5.5 and those with lower magnitudes.

Precision: The ratio of true positive predictions to the total predicted positives. High precision indicates a low false positive rate.
Recall: The ratio of true positive predictions to the total actual positives. High recall indicates a low false negative rate.
F1-score: The harmonic mean of precision and recall, providing a single metric that balances both concerns.

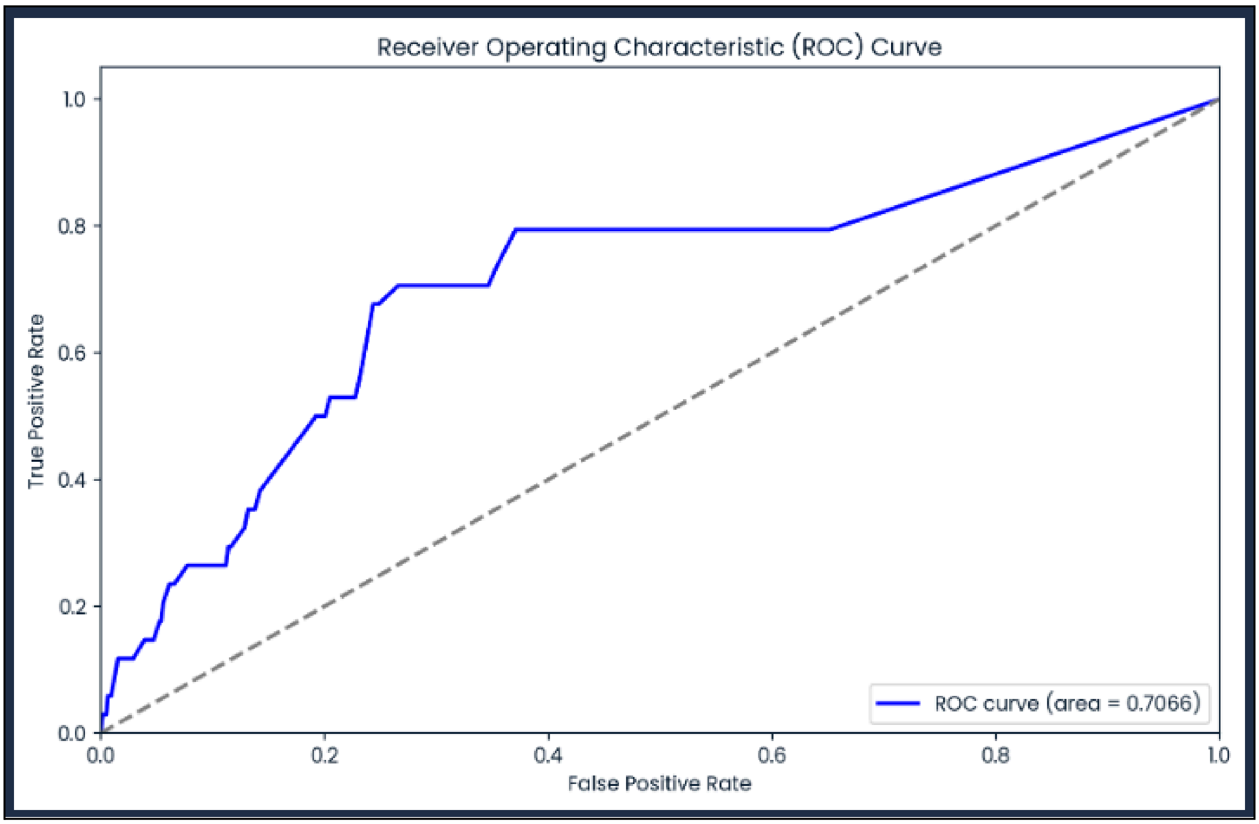

Chart No.5 Shows the ROC (Receiver Operating Characteristic) curve that provides insights into the performance of a model.

Insights:
The ROC curve analysis reveals that the model has a moderate AUC score of approximately 0.7066, indicating some discriminative power but not high accuracy. It illustrates the trade-off between true positive rate (sensitivity) and false positive rate, showing good sensitivity at lower false positive rates but diminishing returns as the false positive rate rises. Overall, the model is reasonably effective but may require further tuning or feature engineering to enhance its predictive accuracy.

An ROC curve is a two-dimensional plot that illus-trates how well a classier system works as the dis-crimination cut-off value is changed over the range of the predictor variable (Yang and Berdine, 2017).

Results and Discussion

The predictor features were defined and analyzed using the ROC AUC score. This study also identified the importance of these features, highlighting which ones should be considered as the main factors influencing the target variable, which in this case is the magnitude. Feature selection is indenspensable stage of data analysis when dealing withdatasets described with thousands variables (Kursa and Rudnicki, 2011).

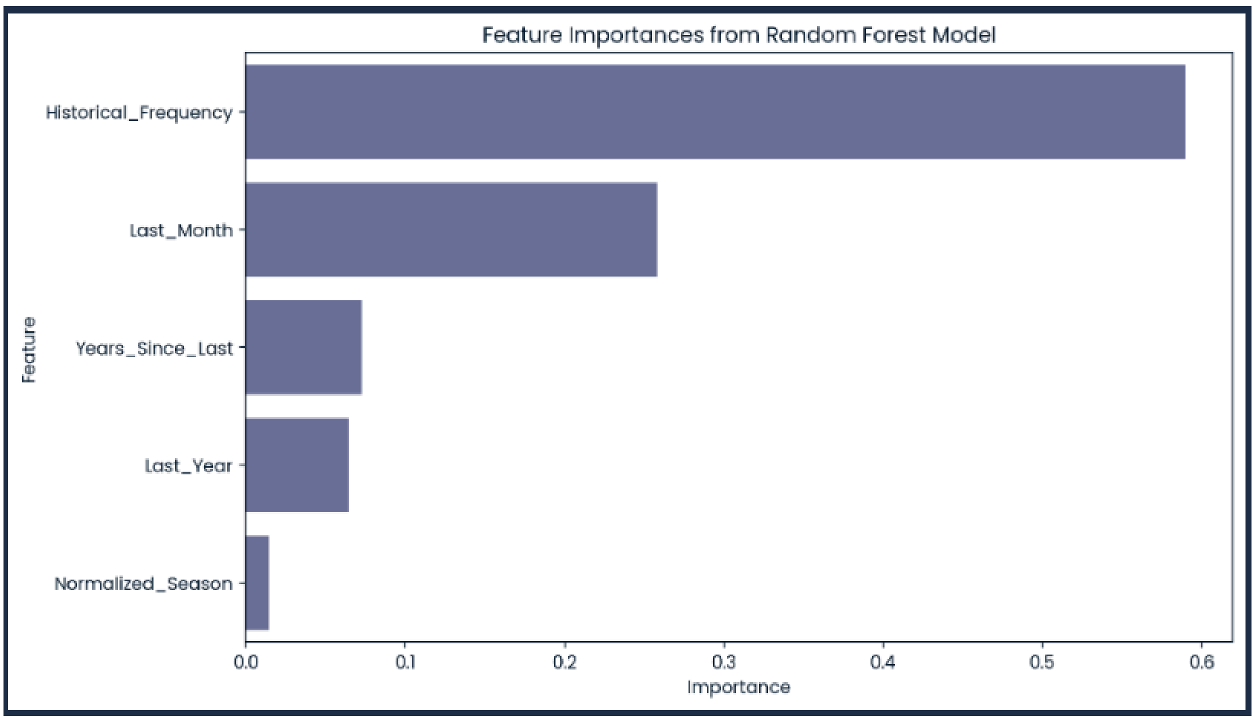

Chart No. 6 Shows the feature importances derived from a Random Forest model.

Insights:
The most influential feature is Historical_Frequency, with an importance of about 0.6, followed by Last_Month at around 0.3. Years_Since_Last and Last_Year also contribute, but to a lesser extent. Normalized_Season has the lowest importance, indicating it has minimal impact on predictions**.**

Time series forecasting algorithms provide information about possible situa-tions in the future, and can be used to anticipate crucial decisions (Leverger et al, 2021).The historical frequency feature was utilized to calculate probabilities for future earthquakes of magnitude 5.5 and above occurring in specific provinces and towns.

Refer to earthquake_prediction_notebook.ipynb for the codes in Python.

Logistic regression is a prediction algorithm used to classify data into categories based on the features of samples (AlShammari, 2024). Earthquake records of magnitude 5.5 or higher is considered in the data prediction model of its next occurrence. Magnitude range between 4.3 to 5.5 can cause damages to buildings and some infrastructures (PHIVOLCS, 2018).

Data modeling and logistic regression prediction steps done:

1. Filtered the town and province earthquake records with magnitude 5.5 or higher.
2. Group by Province and Town, then calculate the mean of relevant features.
3. Prepare the features and target variable for the logistic regression model.
4. Handle missing values by imputing the mean.
5. Standardize the features.
6. Split the data into training and testing sets.
7. Initialize and train the logistic regression model.
8. Predict the probabilities for the test set.
9. Add the forecast probabilities to the dataframe.
10. Round the forecast probabilities to 3 decimal places.
11. Group by Province and calculate the mean forecast probability for each province.
12. Merge the forecast probabilities with the original grouped dataframe to include towns.
13. Save Province Forecast with Towns to csv file named probability_forecast.csv.

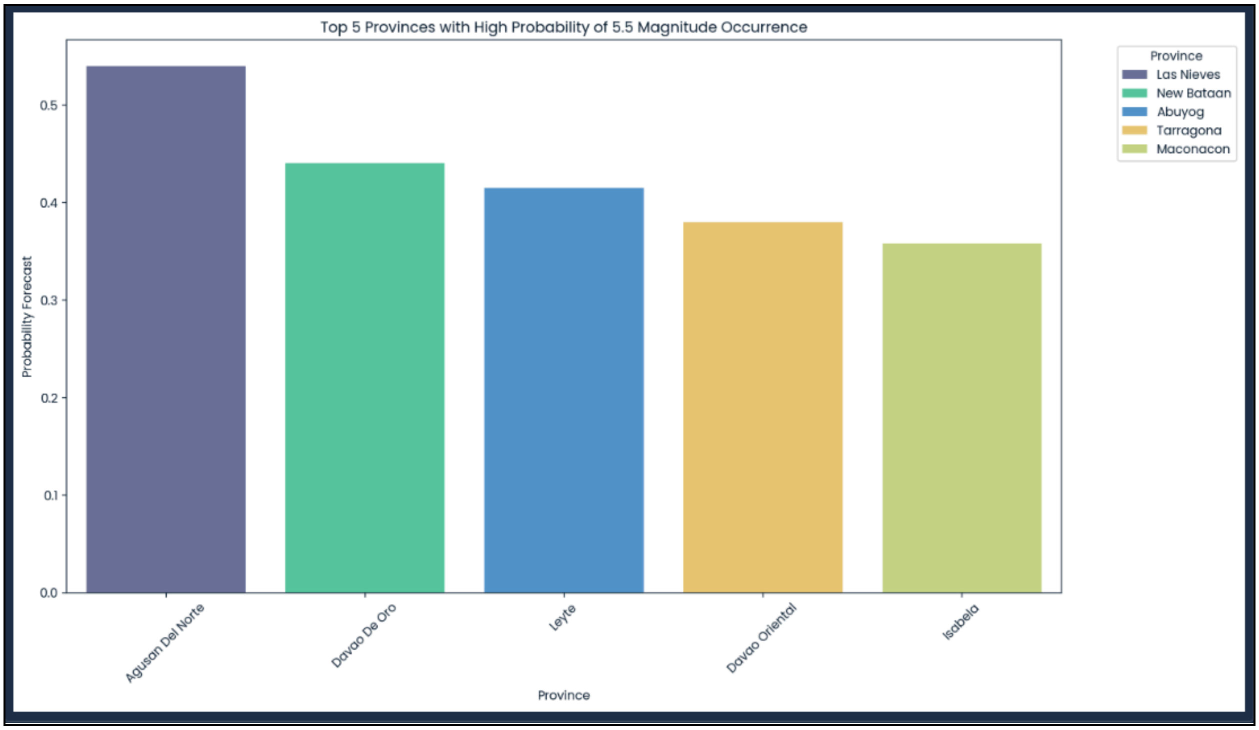

Chart No. 7 Illustrates the top 5 provinces in the Philippines with the highest probability of experiencing a 5.5-magnitude earthquake occurrence.

Insights: Agusan del Norte is at the highest risk for earthquakes, followed closely by Davao de Oro and Leyte. Davao Oriental and Isabela also face notable risks, indicating a widespread threat across various provinces in the Philippines. This information can help local authorities and residents focus on earthquake preparedness in these high-risk areas.

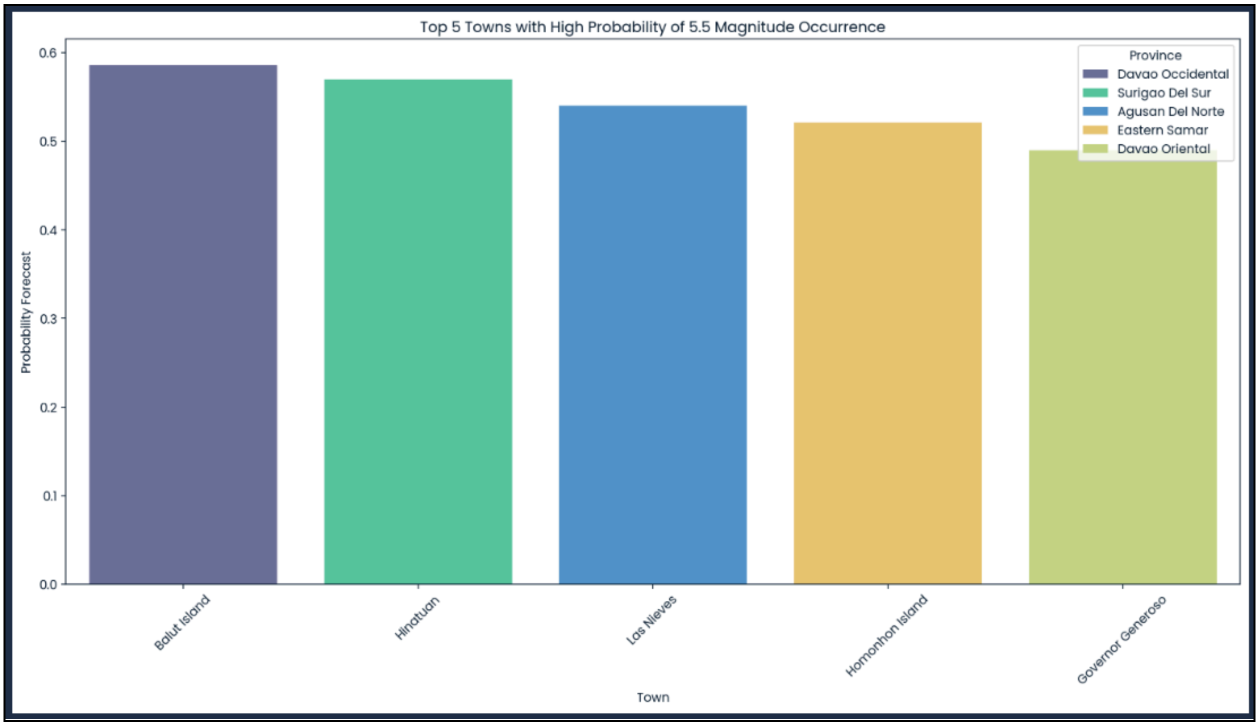

Chart No. 8 Illustrates the top 5 towns in the Philippines with the highest probability of experiencing a 5.5-magnitude earthquake occurrence.

Insights:
Balut Island and Hinatuan in Surigao del Sur have the highest forecasted probabilities for 5.5 magnitude earthquakes, close to 0.6. Las Nieves in Agusan del Norte also shows significant quake probabilities. Homohon Island in Eastern Samar and Governor Generoso in Davao Oriental have slightly lower, yet notable, risks. This data highlights the widespread seismic risks across the Philippines, especially in Surigao del Sur, Agusan del Norte, and Eastern Samar, which is important for local governments in disaster management and informing residents about potential seismic activity.

The aggregated features dataset is referenced for the machine learning model that will predict the magnitude and probability of an earthquake occurring on a projected date and time. Linear regression is a statistical test applied to a data set to define and quantify the relation between the considered variables (Kumari and Yadav, 2018).
Linear regression analysis steps done:

1. Define features and target variables.
2. Drop rows with NaN values in the target variable.
3. Align X with y after dropping NaN values.
4. Split the data into training and testing sets.
5. Define preprocessing for numerical features: impute missing values and scale.

6. Bundle preprocessing for numerical features.
7. Define the model.
8. Create and evaluate the pipeline.
9. Train the model.
10. Predict on the test set.
11. Define metrics.
12. Generate predictions for the next occurrence of earthquakes.
13. Add projected year, month, day, and time.
14. Filter out past dates.
15. Relevant columns for the output file selection.
16. Forecast live the next occurrence.
17. Add magnitude level to the predictions output.
18. Add a probability column predicting the certainty of the predicted magnitude to occur in the future.
19. Convert probability to percentage with 2 decimal places.
20. Save the predictions to a new csv file named as earthquake_predictions.csv.

Statistical metrics are then measured to check model performance.

1. Mean Squared Error (MSE), the average squared difference between the value observed in a statistical study and the values predicted from a model (Ken, 2024).
2. R-squared (R2) is interpreted as representing the percentage of variation in the dependent variable explained by variation in the independent variables (Figueiredo et al, 2011).
3. Mean Absolute Error (MAE) is a regressive loss measure looking at the absolute value difference between a model’s predictions and ground truth, averaged out across the dataset (Burch, 2023).
4. Explained variance (sometimes called “explained variation”) refers to the variance in the response variable in a model that can be explained by the predictor variable(s) in the model (Bobbitt, 2022).

```
Mean Squared Error (MSE): 7.301277521815001e-31
R-squared (R2): 1.0
Mean Absolute Error (MAE): 6.433799607076993e-16
Explained Variance Score (EVS): 1.0
```

Image No. 1 Shows the metrics considered to measure the performance of the model.

Insights:

1. **High Accuracy and Predictive Strength:**
   1.1 The R-squared (R^2) value of 1.0 signifies that the model explains 100% of the variance in the target variable, indicating perfect predictions on the dataset used for evaluation.
   1.2 The Explained Variance Score (EVS) of 1.0 further reinforces that the model consistently captures variability without error.
2. **Minimal Error:**
   2.1 The Mean Squared Error (MSE) is extremely close to zero (7.30e-31), showing that the squared differences between predicted and actual values are negligible.
   2.2 The Mean Absolute Error (MAE), also near zero (6.43e-16), confirms that predictions deviate from true values by an insignificant margin.
3. **Model Robustness:**
   3.1 These performance metrics suggest the model is highly reliable, making it well-suited for applications in earthquake forecasting based on historical data trends.

Conclusions

The insights presented below have been gathered from this study:

1. There is a strong link between high historical earthquake frequency and larger magnitudes.
2. The ROC curve analysis reveals a moderate AUC score of approximately 0.7066, suggesting some predictive capability.
3. The key feature influencing earthquake predictions is Historical_Frequency.
4. There is a widespread seismic risks across the Philippines, especially in Surigao del Sur, Agusan del Norte, and Eastern Samar

Overall, the model effectively explains a significant portion of variance and performs well, as shown by its low mean squared error.

Recommendation

Based on the insights gathered from the study, here are several key recommendations for addressing earthquake risk in the Philippines:

1. Prioritize Earthquake Preparedness in High-Risk Areas especially in terms of budget allocation.
2. Enhance Predictive Models Using Historical Data.
3. Refine Predictive Capabilities with Machine Learning.
4. Strengthen Infrastructure in Vulnerable Regions.
5. Develop Comprehensive Regional Risk Maps.
6. Public Awareness Campaigns, Information Drive and Trainings.

References

U.S. Geological Survey. *Earthquake Introduction*. U.S. Department of the Interior, 30 Nov. 2016, http://pubsdata.usgs.gov/pubs/gip/earthq1/intro.html

Adagunodo, T.A., and L.A. Sunmonu. "Earthquake: A Terrifying of All Natural Phenomena." *Journal of Advances in Biological and Basic Research*, vol. 1, 2015, pp. 4-11.https://www.researchgate.net/publication/303910512_Earthquake_a_Terrifying_of_all_Natural_Phenomena.

Philippine Institute of Volcanology and Seismology (PHIVOLCS). "Introduction to Earthquake: What Is an Earthquake?" *PHIVOLCS*, 2018, https://www.phivolcs.dost.gov.ph/index.php/earthquake/introduction-to-earthquake.

United Nations Office for Disaster Risk Reduction (UNDRR). *Disaster Risk Reduction in the Philippines*: *Status Report 2019*. Regional Office for Asia and the Pacific, 2019. Bangkok, Thailand, https://www.unisdr.org/files/68265_682308philippinesdrmstatusreport.pdf.

Rusydy, Ibnu, et al. "A GIS-Based Earthquake Damage Prediction in Different Earthquake Models: A Case Study at the University of the Philippines Los Baños, Philippines." *Philippine Journal of Science*, vol. 147, no. 2, June 2018, pp. 301-316. ISSN 0031-7683. Received 11 Aug. 2017. https://www.researchgate.net/publication/325567388_A_GIS-Based_Earthquake_Damage_Prediction_in_Different_Earthquake_Models_A_Case_Study_at_the_University_of_the_Philippines_Los_Banos_Philippines.

Lagmay, Alfredo Mahar Francisco A., et al. "Disseminating Near-Real-Time Hazards Information and Flood Maps in the Philippines Through Web-GIS." *Journal of Environmental Sciences*, vol. 59, 2017, pp. 13-23. https://doi.org/10.1016/j.jes.2017.03.014.

Bollettino, Vincenzo, et al. "Public Perception of Climate Change and Disaster Preparedness: Evidence from the Philippines." *Climate Risk Management*, vol. 30, 2020, p. 100250. https://doi.org/10.1016/j.crm.2020.100250.

Mejia, Jacenth, et al. "Earthquake Prediction through Kannan-Mathematical-Model Analysis and Dobrovolsky-Based Clustering Technique." *DLSU Research Congress 2015*, De La Salle University, Manila, Philippines, 2-4 Mar. 2015. Physics Department, De La Salle University.

Saabith, A. L. Sayeth, MMM. Fareez, and T. Vinothraj. "Python Current Trend Applications - An Overview: Popular Web Development Frameworks in Python." *International Journal of Advance Engineering and Research Development*, vol. 6, no. 10, Oct. 2019, e-ISSN: 2348-4470, p-ISSN: 2348-6406. https://www.researchgate.net/publication/373633075_The_Rise_of_Python_A_Survey_of_Recent_Research.

BwandoWando. *Philippine Earthquakes (from PHIVOLCS) [Data set].* Kaggle, 2024, https://doi.org/10.34740/KAGGLE/DS/5555087.

Singh, Nilu. "Data Cleaning Methods." *Koneru Lakshmaiah Education Foundation (K.L.E.F.), Department of Computer Science & Engineering*, May 2023, https://www.researchgate.net/publication/370496937_Data_Cleaning_Methods.

Murel, Jacob, and Eda Kavlakoglu. "What Is Feature Engineering?" *IBM*, 20 Jan. 2024, https://www.ibm.com/topics/feature-engineering.

Camizuli, Estelle, and Emmanuel John M. Carranza. "Exploratory Data Analysis (EDA)." *November 2018*. DOI:10.1002/9781119188230.saseas0271. https://www.researchgate.net/publication/329204518_Exploratory_Data_Analysis_EDA.

Cao, Shengjia, Yunhan Zeng, Shangru Yang, and Songlin Cao. "Research on Python Data Visualization Technology." *Journal of Physics: Conference Series*, vol. 1757, no. 1, 2021,DOI:10.1088/1742-6596/1757/1/012122.https://www.researchgate.net/publication/349009998_Research_on_Python_Data_Visualization_Technology.

Kostyuchenko, Yulia, and Ilya Gosudarev. "Analysis of Approaches to Data Modeling Using Python Libraries." *ITMO University Scientific and Educational Conference*, Feb. 2018, Saint Petersburg, Russia. https://www.researchgate.net/publication/335600632_Analysis_of_approaches_to_data_modeling_using_Python_libraries.

Asefi, Sajjad, et al. "Anomaly Detection, Classification and Identification Tool (ADCIT)." *Software Impacts*, vol. 16, 2023, p. 100465. https://doi.org/10.1016/j.simpa.2023.100465.

Boussiala, Mohamed Nachid. *Correlation Analysis in Python.* November 2021. Algiers University. DOI:10.13140/RG.2.2.14628.40326. GitHub Repository, https://github.com/M-nachid.

Ali, Peshawa Jammal Muhammad, and Rezhna Hassan Faraj. *Data Normalization and Standardization: A Technical Report.* January 2014. Koya University. DOI:10.13140/RG.2.2.28948.04489.

Ali, Jehad, et al. "Random Forests and Decision Trees." *ResearchGate*, Sept. 2012, https://www.researchgate.net/publication/259235118_Random_Forests_and_Decision_Trees.

Kursa, Miron B., and Witold Rudnicki. "*The All Relevant Feature Selection Using Random Forest.*" June 2011. https://www.researchgate.net/publication/51913500_The_All_Relevant_Feature_Selection_using_Random_Forest.

Yang, Shengping, and Gilbert Berdine. "The Receiver Operating Characteristic (ROC) Curve." *The Southwest Respiratory and Critical Care Chronicles*, vol. 5, no. 19, May 2017, pp. 34. Texas Tech University Health Sciences Center, doi:10.12746/swrccc.v5i19.391. https://www.researchgate.net/publication/316751328_The_receiver_operating_characteristic_ROC_curve.

AlShammari, Ahmad Farhan. "Implementation of Logistic Regression Using Gradient Descent in Python." *International Journal of Computer Applications*, vol. 186, no. 13, Mar. 2024, pp. 41–46. DOI: 10.5120/ijca2024923509. https://www.researchgate.net/publication/383264729_Implementation_of_Logistic_Regression_using_Gradient_Descent_in_Python

Kumari, Khushbu, and Suniti Yadav. "Linear Regression Analysis Study." *Journal of the Practice of Cardiovascular Sciences*, vol. 4, no. 1, Jan.–Apr. 2018, pp. 33–36. DOI: 10.4103/jpcs.jpcs_8_18. https://journals.lww.com/jpcs/fulltext/2018/04010/linear_regression_analysis_study.9.aspx.

Stewart, Ken. "Mean Squared Error." *Mathematics*, Encyclopædia Britannica, 24 Sept. 2024, https://www.britannica.com/science/mean-squared-error.

S. Kassam, "The mean-absolute-error criterion for quantization," ICASSP '77. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Hartford, CT, USA, 1977, pp. 632-635, doi: 10.1109/ICASSP.1977.1170242.

Burch, David. "Mean Absolute Error in Machine Learning: What You Need to Know." *Published*, 9 Sept. 2023. https://arize.com/blog-course/mean-absolute-error-in-machine-learning-what-you-need-to-know/.

Bobbitt, Zach. "What Is Explained Variance? (Definition & Example)." *Statology*, 20 June 2022, https://www.statology.org/explained-variance/.

Philippine Institute of Volcanology and Seismology. "Primer on the 06 July 2017 Magnitude 6.5 Leyte Earthquake." *Philippine Institute of Volcanology and Seismology*, 7 July 2017, https://www.phivolcs.dost.gov.ph/index.php/news/631-primer-on-the-06-july-2017-magnitude-6-5-leyte-earthquake-07-july-2018.