RESEARCH ARTICLE OPEN ACCESS

Soil Liquefaction Susceptibility Mapping Using a Machine Learning Algorithm: A Case of Guagua, Pampanga

Mark Benedic E. Dumantay[1],Juliana Marie C. Abigan[2],Trixie B. Dela Cruz[3], Mikaella E. Feliciano[4], Reigner C. Gigante[5], Ma. Angelu S. Castro, RCE, RMP[6], Charles G. Lim, ECE, ECT[7]

[1]Department of Civil Engineering, Don Honorio Ventura State University, Villa de Bacolor, Pampanga, Philippines
Email: dumantay.markbenedic.8sscb@gmail.com

[2]Department of Civil Engineering, Don Honorio Ventura State University, Villa de Bacolor, Pampanga, Philippines
Email: abiganjulianamarie15@gmail.com

[3]Department of Civil Engineering, Don Honorio Ventura State University, Villa de Bacolor, Pampanga, Philippines
Email: trixiedelacruz23@gmail.com

[4]Department of Civil Engineering, Don Honorio Ventura State University, Villa de Bacolor, Pampanga, Philippines
Email: mikaellafeliciano20@gmail.com

[5]Department of Civil Engineering, Don Honorio Ventura State University, Villa de Bacolor, Pampanga, Philippines
Email: reignergigante08@gmail.com

[6]Department of Civil Engineering, Don Honorio Ventura State University, Villa de Bacolor, Pampanga, Philippines
Email: mscastro@dhvsu.edu.ph

[7]Department of Civil Engineering, Don Honorio Ventura State University, Villa de Bacolor, Pampanga, Philippines
Email: cglim@dhvsu.edu.ph

--************************----------------------------------

Abstract:

This case study investigates the application of a machine learning algorithm in predicting the soil liquefaction susceptibility of Guagua, Pampanga. The random forest was employed in training the predictive model using 413 datasets. The variables in the dataset include natural moisture content, D50, plasticity index, SPT N-Value, groundwater level, and fines content. Ethical considerations ensured the privacy of the data collected from various soil testing centers and DPWH districts in Pampanga. The trained model achieved 93% accuracy in predicting soil liquefaction susceptibility after feeding it with the datasets. In addition, the feature importance revealed that variables such as D50 and groundwater level have a more significant influence on the likelihood of soil liquefaction compared to fines content, SPT N-Value, PI, and natural moisture content. Based on the model, the locale has 16 barangays that are susceptible to soil liquefaction, 3 barangays that are not susceptible to soil liquefaction, and 12 barangays with unidentified soil liquefaction susceptibility. The researchers had concluded that Guagua, Pampanga has a high probability of experiencing soil liquefaction based on the hazard map generated from a web application, equipped with the predictive model trained using a random forest algorithm, developed by the researchers.

Keywords **—soil liquefaction susceptibility, random forest, standard penetration test, hazard map**

--************************----------------------------------

I. INTRODUCTION

Over the years, earthquake-induced damage has been continually heightened by liquefaction. It is the loss of contact between soil particles during earthquakes. This phenomenon typically occurs in saturated, loose, sandy soils where there is not enough time for water to drain out of the pores, resulting in higher excess pore pressure and causing sand particles to float. Furthermore, liquefaction causes a variety of ground failures, such as loss of bearing capacity, lateral spreading, and flow, ultimately resulting in building collapses[1], [2], [3]. Thus, there is a need for detailed liquefaction analysis in the form of a hazard map that will help reduce liquefaction-related incidents in future construction projects. Soil liquefaction potential depends on the geotechnical properties of the soil and the groundwater table level, and can be triggered by earthquakes and volcanoes due to nearby fault lines.

Factors such as soil characteristics and the groundwater table level influence the likelihood of liquefaction to occur. In geotechnical engineering, 'soft sediment' describes a land mass or formation with high water content[4]. Regions with high soil strength generally have low liquefaction susceptibility, while areas with low soil strength are more prone to liquefaction[5]. In addition to soil properties, the groundwater table level contributes to the saturation of the soil, affecting the liquefaction potential [6]. Finally, soil liquefaction is a secondary effect of earthquakes and, in some cases, can be induced by heavy rains.

Liquefaction is a phenomenon experienced in different parts of the world. In some cases, the effects of soil liquefaction lead to the collapse of several structures, resulting in the loss of human lives [7]. One notable incident of soil liquefaction occurred in San Francisco, California, where an earthquake-induced liquefaction caused the devastation of the Bay Bridge, an apartment, and an entire street of the said city [8]. Another case occurred in Palu, Indonesia, where 1,747 homes were destroyed in the neighborhood [9]. Additionally, a study documenting the effect of larger liquefaction events since 1900 listed numerous fatalities and economic issues, with death counts reaching up to 16 persons [10].

The Philippines is situated in the Pacific Ring of Fire, where 91% of the world's earthquakes and volcanic activities occur. As previously mentioned, liquefaction is a secondary effect of earthquakes, making the Philippines also prone to soil liquefaction. The provinces of the Philippines have recorded significant effects of liquefaction. In Dagupan, where the water table is shallow, reinforced concrete buildings settled due to liquefaction, and severe tilts were observed, measuring up to 50–75 cm and 1-2.5 degrees, respectively[11]. In Davao Del Sur, located in the lower part of the Philippines, the Mindanao; a three-story commercial building collapsed due to liquefaction triggered by a magnitude 6.9 earthquake, and three (3) confirmed deaths were recorded [12].

Going deeper, Central Luzon is where most seismic and volcanic activities are active. The seismic activities in this area are generally caused by different faults, namely, the Philippine Fault, Iba Fault, West Valley Fault System, East Zambales Fault, and other known faults in the locality. With that, the Central Luzon, where the province of Pampanga is part of, is just as at risk of soil liquefaction [13]. According to the Philippine Institute of Volcanology and Seismology (PHIVOLCS) former director, Renato Solidum Jr., the province of Pampanga is vulnerable to strong shaking and softer soil due to the deposited lahar in different parts of the province after the eruption of Mt. Pinatubo in 1991. These areas include Guagua, Floridablanca, and the City of Angeles [14]. Due to ground subsidence, a tower sank and remained tilted up to this day, it is known as the 'Leaning Water Tower of San Fernando, Pampanga [15].

The Department of Public Works and Highways (DPWH) has been long preparing for the 'Big One' by building action centers in the different parts of Region III [16]. The 'Big One' is one of the most anticipated earthquake events that could potentially devastate the Philippines [17]. The 'Big One' is expected to have a magnitude of at least 7.2, and it is characterized as a "Very Destructive" Intensity VIII earthquake according to the PHIVOLCS Earthquake Intensity Scale [18]. This speculated Intensity VIII earthquake can cause buildings and infrastructures to settle, topple, and to be destroyed. In addition, Intensity VIII earthquakes can cause lateral spreading and liquefaction.

According to the FaultFinder of the Department of Science and Technology (DOST) [19], the nearest active faults in Guagua, Pampanga are the Iba fault and the Valley Fault Systems, the distance range from 43-54 km. Since Guagua is situated between two active faults, this generally means that the city is prone to earthquakes that may induce liquefaction. Moreover, Guagua often experiences flooding due to its geographic location, and that contributes to soil becoming softer. These evidences prove the need to procure a liquefaction hazard map in Guagua, Pampanga, all for the safety and awareness of the residents of the city.

The study aimed to encompass the gap in producing a soil liquefaction susceptibility map using a machine learning algorithm. While several studies had gathered data for liquefaction hazard maps and analyzed it using data analytics, machine learning algorithms were not utilized. On the other hand, there were multiple studies regarding the assessment of soil liquefaction; however, no liquefaction hazard maps were being produced. In the construction of structures, many civil engineers exclude information about soil liquefaction due to its low availability. This exclusion can lead to incidents where buildings cannot withstand the hazard posed by soil liquefaction. Therefore, the researchers intended to produce a hazard map using a machine learning algorithm, random forest, incorporated in a web application for the Municipality of Guagua, Pampanga.

II. METHODS

In this chapter, the methodological framework was discussed along with the research design, system design, research locale, research instrument, and the data collection methods that were used for data analysis.

A. *Research Locale*

The study was conducted in Guagua, Pampanga, where the municipality has a land area of 48.67 square kilometers. Guagua was formerly named "Wawa," which means river mouth, and is located along a river [20]. This municipality consists of 31 barangays in total, and based on the 2020 census, it has a total population of 128, 893 [21]. The study mainly focused on the possible occurrence of soil liquefaction in the 31 barangays of Guagua, Pampanga. Given that the said area is located along a river, it is also composed of fine sand, silty loam, and hydrosol, and it is alluvial. Alluvial soil type has weak and shallow profiles, and it is considered immature soil since it is incomplete in its soil profile [20]. Since the municipality is ranked first and has a fast-growing economy, it is vital to have a hazard map for the said locale [22].

B. *Research Instrument*

To obtain the preliminary data, the researchers prepared a formal letter requesting for standard penetration test and borehole test data that was addressed to the DPWH 2nd District Engineering Office, the DPWH Regional Office, and the Unified Geotest Laboratory. Also, the letter was signed by the thesis coordinator, adviser, and also by the chairperson. Following the approval from the said office personnel, the collected data was recorded through Microsoft Excel.

C. *DataCollectionMethod*

A request letter to conduct the study was prepared and signed by the research coordinator, adviser, and also by the chairperson, and the researchers for the collection of data from the DPWH 2nd District Engineering Office, the DPWH Regional Office, and the Unified Geotest Laboratory concerning the standard penetration tests and borehole tests.

The study was conducted in Guagua, Pampanga, in reason that the area shows huge potential for liquefaction. Upon the acceptance of the request letter, the data collected was tabulated and organized through Microsoft Excel. The recorded data was interpreted using the Random Forest Algorithm.

D. Data Analysis

The following statistical method was used by the researchers to analyze the data that was gathered from the DPWH 2nd District Engineering Office, the DPWH Regional Office, and the Unified Geotest Laboratory:

1) Machine Learning: Machine learning algorithms use data to create predictions by constructing mathematical models from multiple datasets that include data from training, validating, and test sets. Factors such as training, validation, and test datasets are crucial for developing reliable and accurate models [23]. The evaluation for liquefaction potential can branch out to multiple problems. Researchers utilized many machine learning approaches for the liquefaction potential assessments to overcome the aforementioned issue. There are only a few studies that apply machine learning models and techniques to predict the probability of liquefaction.

2) Random Forest Algorithm: The random forest (RF) is a type of algorithm that creates numerous predictors through ensemble learning that is based on statistical theory, which can tackle classification and regression using classification trees and regression trees, respectively [24]. The advantage of utilizing random forest lies in its simplified hyperparameter selection and its ability to address overfitting issues. This method was proven effective in various geotechnical engineering issues. However, few studies discussed the application of random forest models in liquefaction assessment[25].

3) Input and Output Variable: As shown in Table I: Input and Output Variables, seven (7) variables were considered in this study. The considered variables were N-Value (number of blow counts), PI (plasticity index), w (natural water content), D50 (average grain size), GWL (groundwater level), and FC (fines content) which were inputted variables to be used in determining the susceptibility of a certain area to liquefaction. And the last variable, soil liquefaction, was the target or output variable of the study.

TABLE I.
INPUT AND OUTPUT VARIABLES

Abbreviation	Description	Units
N-Value	Standard Penetration Test N-Value	--
PI	Plasticity Index	Percent (%)
ω	Natural Water Content	Percent (%)
D_{50}	Mean Particle Size	mm
FC	Fines Content	Percent (%)
GWL	Ground Water Level	m
Liq	Liquefaction	--

4) Programming (Backend): Python is one of the multi-purpose, open-source, and advancing programming languages, and can be used both in web and software applications. Furthermore, python programming can be used in Data manipulation and visualization, Statistics, Mathematics, and Machine Learning, and in many more applications [26]. The term data analytics is described as the process of analyzing numerous data sets to make predictions and decision-making with sufficient accuracy and precision. As this study focused on predictions and regression, the researchers used the Sci-kit Learn as a library for Python programming as this library is best suited for regression analysis, classification, and model clustering [27]. Also, the researchers utilized JavaScript Object Notation (JSON), a file format suitable for human-readable language. JSON is mostly used in web applications, fitting for the objective of this study. Lastly, the researchers used the Flask API, just like JSON and Python, it is also an ideal choice for web applications [28].

5) Mapping (Frontend): The researchers utilized GeoJSON for the mapping of the locale, Guagua, Pampanga. The researchers used the GeoJSON to get the coordinates of each barangay of Guagua, Pampanga to easily map out the soil liquefaction susceptibility of Guagua, Pampanga. The GeoJSON is powered by Mapbox, the coordinates were downloaded from their website and were incorporated into the web application developed by the researchers with the help of a programmer[29]. In addition, Google Maps will be integrated into the web application programming interface (API) as the main source for printing the result of susceptibility mapping. Also, Bootstrap was used alongside GeoJSON and Google Maps as one of the frontends of the web application. Bootstrap is a frontend framework designed to be an open source for web applications that is capable of providing responsive templates that can automatically adapt display in response to the source code. All these three comprise the frontend of the web application of the researchers.

6) Performance Indicator:The confusion matrix is a classification model used in binary problems there are only two classes, in this case, soil liquefaction susceptible and not susceptible [30].]. In this study, 0 and 1 were used to represent "not susceptible" to liquefaction and "susceptible" to liquefaction accordingly. The confusion matrix has four possible predictions as shown in the 2 by 2 contingency table in Figure 1. The green diagonals represent the correct predictions and the red diagonals represent the incorrect predictions.

These are the possible outcomes of the machine learning model:

True Negative (TN): Correct predictions of not susceptible to soil liquefaction.

False Negative (FN): Incorrect predictions of not susceptible to soil liquefaction.

False Positive (FP): Incorrect predictions of susceptible to soil liquefaction.

True Positive (TP): Correct predictions of susceptible to soil liquefaction.

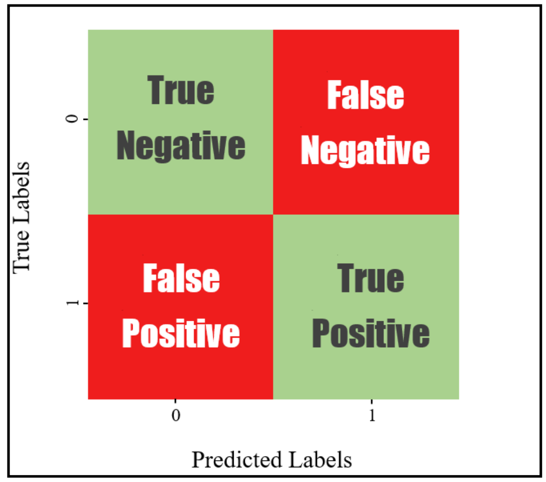


Fig. 1 Confusion Matrix of the Study

Accuracy

Accuracy is the ratio of all of the correct predictions to all of the predictions made. In simpler terms, accuracy shows how often the model makes correct predictions [31]. The accuracy can be calculated by Equation 1.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$
(Eq. 1)

Precision

Precision is the ratio of all correct positive or negative predictions to all of the positive or negative predictions made by the model. Simply, this means that precision determines how often the model is correct when predicting a chosen class [31]. The precision can be calculated by the Equations 2 (Positive) and 3 (Negative).

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad \text{(Eq. 2)}$$

$$Precision = \frac{TN}{TN+FN} \times 100\% \quad \text{(Eq. 3)}$$

Recall

Recall shows the true predictions made for all of the positive or negative datasets. In other words, it is the ratio of true predictions to a target dataset [31]. Recall can be calculated as shown in Equations 4 (Positive) and 5 (Negative).

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad \text{(Eq. 4)}$$

$$Recall = \frac{TN}{TN+FP} \times 100\% \quad \text{(Eq. 5)}$$

F1-score

The F1-score can be understood as the balanced average of precision and recall as presented in Equation 6. This can be calculated as the same as calculating harmonic mean. The closer the value of f1-score to 1 implies that the model has high precision and recall [32].

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad \text{(Eq. 6)}$$

As previously mentioned, the gathered data was split into two parts: the training data and the test data. The data was cut into a 75% to 25% ratio, with 75% of the data being the training data and the remaining 25% of the data being used as the test data [33]. To determine the performance of the machine learning model, the test data was used as inputs in the web application and its score will be measured. Four performance metrics were utilized in the study, namely: F1 score, accuracy, precision, and recall. These four performance metrics evaluated the score of the machine learning model trained using the random forest algorithm using the test data. These performance metrics are commonly used for binary classification problems, hence, they were useful in predicting if a soil is susceptible to liquefaction or not.

The feature importance, in the context of machine learning, indicates how each variables contributes to the ability of the machine learning model to make correct and incorrect predictions[34]. Feature importance is generally used in decision tree algorithms such as gradient boosting, and the algorithm used in this study, random forest. Since feature importance measures the degree on how each variable contributes to the predictions, it also has the function to rank the variables presented in this study. The ranking starts with the best splitting factor to the least important variable.

III. RESULTS AND DISCUSSION

This chapter consists of the data analysis, results, and discussion of the findings. Additionally, the results of the study are based on the problem statements and research objectives that focus on producing a Soil Liquefaction Susceptibility Map in Guagua, Pampanga.

A. *Results*

After the collection of SPT data, the researchers proceeded in organizing and sorting the SPT data into datasets. The summary of data is presented below, as shown in Table II. Table II shows the detailed statistical description of the datasets with a count of 413. The SPT N-value ranges from 1 to 39.5, the plasticity index varies between 0 to 45.49%, the natural moisture content ranges between 0 to 95.3%, the average grain size measures from 0 to 2 mm, the fines content varies between 0 to 97.77%, and the groundwater level measures at depth of 0.5 m to 36 m below natural grade line. Furthermore, the variation and distribution of each variable containing 413 data is also shown in Figure 2.

TABLE II
THE SUMMARY OF THE SPT DATASETS

	N	IP (%)	ω (%)	D50 (mm)	FC (%)	Gwl (m)	Liq
Count	413.00	413.00	413.00	413.00	413.00	413.00	413.00
Mean	12.37	5.54	21.46	0.37	33.57	6.12	0.82
STD	9.07	11.14	10.24	0.33	25.77	5.15	0.38
Min	1.00	0.00	2.73	0.14	1.07	0.50	0.00
25%	5.33	0.00	15.04	0.17	16.00	3.00	1.00
50%	9.50	0.00	18.43	0.20	25.62	5.00	1.00
75%	18.00	6.00	24.87	0.40	40.88	9.00	1.00
Max	39.50	45.49	95.30	2.00	97.77	36.00	1.00

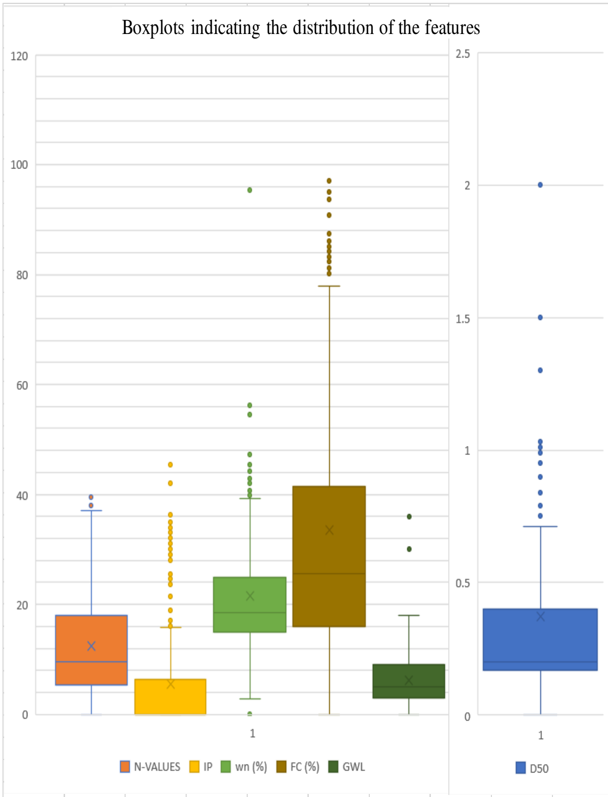


Fig. 2 The Distribution of each Variable

The first objective of this study was to train a machine learning model using random forest algorithm. To achieve this, the collected datasets were fed to the machine learning model and the results of the training were analyzed by the researchers. In order to prove that the researchers developed a working model, the confusion matrix, the performance indicators, the cross-validation, and the feature importances are to be discussed below.

Confusion Matrix

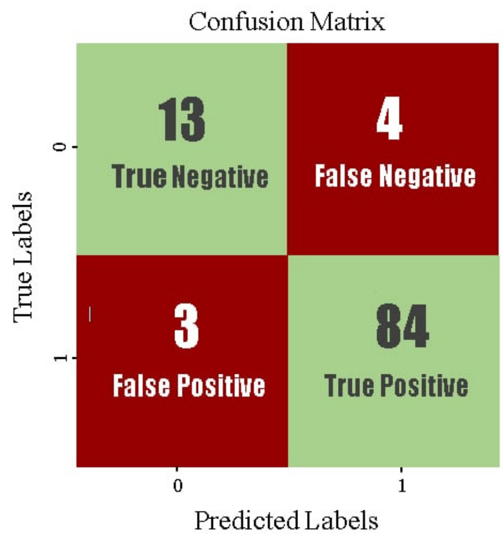


Fig 3 Confusion Matrix Result

The confusion matrix of results, presented in Figure 3, compares the true labels—from the original dataset—to the predicted labels—predictions generated by the model. From the figure, it can be seen that the model generated 84 true positive (TP) predictions, simply, this means that the model correctly predicted 84 datasets that are actually susceptible to liquefaction. However, the model made 4 false negative (FN) predictions, meaning, 4 incorrect predictions were made on datasets that are actually susceptible to liquefaction. On the other hand, the machine learning model was able to produce 13 true negative (TN) predictions and 3 false positive (FP) predictions, 13 correct predictions and 3 incorrect predictions were made in datasets that were not susceptible to liquefaction. The result is summarized below:

TP = 84
FN = 4
TN = 13
FP = 3

Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Accuracy = \frac{84 + 13}{84 + 13 + 3 + 4} \times 100\%$$

$$Accuracy = 93.26923\%$$

The accuracy of the machine learning model is calculated as the ratio of the sum of all correct predictions to the sum of all predictions made by the model as shown in Equation 1. The model garnered an accuracy of 93% which implies that the model has a high chance to make accurate predictions on determining the susceptibility to soil liquefaction.

Positive Predictions

The high metrics score of positive predictions on precision (96.55172%), recall (95.45455%), and f1-score (96%) correspond to a high true positive rate as shown below in the equations. The computations of precision, recall, and f1-score of positive predictions (susceptible to soil liquefaction) are presented below in a consecutive manner.

Precision

$$Precision = \frac{TP}{TP + FP} \times 100\%$$
$$Precision = \frac{84}{84 + 3} \times 100\%$$

$$Precision = 96.55172\%$$

Recall

$$Recall = \frac{TP}{TP + FN} \times 100\%$$
$$Recall = \frac{84}{84 + 4} \times 100\%$$

$$Recall = 95.45455\%$$

F1-score

$$F1 = 2 \times \frac{Precision \ \times Recall}{Precision + Recall}$$
$$F1 = 2 \times \frac{96.55172\% \times 95.45455\%}{96.55172\% + 95.45455\%}$$

$$F1 = 96\%$$

Negative Predictions

Upon investigation, the model has shown an inability to make accurate predictions for those barangays that are not susceptible to liquefaction. This was suggested by the overall score of negative predictions shown below in the equations. For precision, the score is 76.47059%. For recall, it scored an 81.25%. And for the f1-score, it garnered a 78.78788%. The computation of precision, recall, and f1-score of negative predictions (no soil liquefaction) are presented below in a consecutive manner.

Precision

$$Precision = \frac{TN}{TN + FN} \times 100\%$$
$$Precision = \frac{13}{13 + 4} \times 100$$

$$Precision = 76.47059\%$$

Recall

$$Recall = \frac{TN}{TN + FP} \times 100\%$$
$$Recall = \frac{13}{13 + 3} \times 100\%$$
$$Recall = 81.25\%$$

F1-score

$$F1 = 2 \times \frac{Precision \ \times Recall}{Precision + Recall}$$
$$F1 = 2 \times \frac{76.47059\% \times 81.25\%}{76.47059\% + 81.25\%}$$

$$F1 = 78.78788\%$$

Flask Extensions, Session Handling, and Cross-Origin Resource Sharing (CORS) are widely utilized in programming to solve common server-related challenges. Additionally, it enhances security and storage for user-specified data across multiple requests. The programming was done on a laptop with Intel(R) Core(TM) i3-8130U CPU @ 2.20GHz, 2208, and 8 GB RAM in an x64-based PC. The dataset examined in the study is trained with 10-fold cross-validation. Figure 4 presents the accuracy of these cross-validation results, categorized based on various parameters.

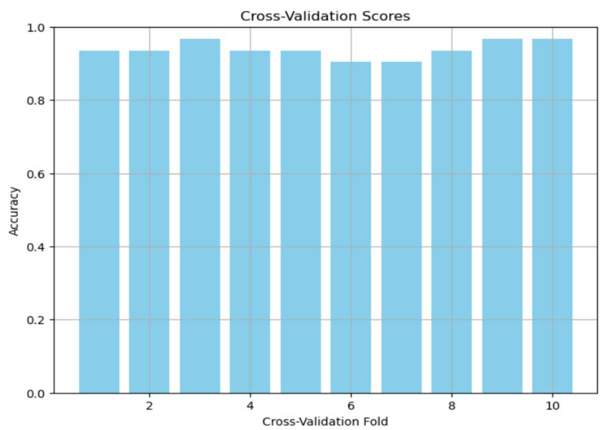


Fig. 4Accuracy of Cross-Validation

TABLE III
TEN OF THE BEST COMBINATIONS

	N	PI (%)	ω (%)	D50 (mm)	FC (%)	Gwl (m)	Liq
0	8.5	32.08	47.34	0.19	80.14	3.10	1
1	5.0	32.09	43.10	0.17	83.14	4.05	1
2	8.0	32.01	42.11	0.17	81.22	3.20	1
3	2.5	0.00	22.71	0.17	30.72	1.68	1
4	1.5	0.00	16.41	0.16	3.66	1.71	1
5	2.0	0.00	17.18	0.17	29.89	1.68	1
6	7.0	0.00	21.24	0.18	32.27	6.00	1
7	8.0	0.00	19.48	0.18	26.80	6.50	1
8	8.0	0.00	22.92	0.16	34.96	7.00	1
9	2.5	0.00	12.83	0.32	7.86	7.18	1

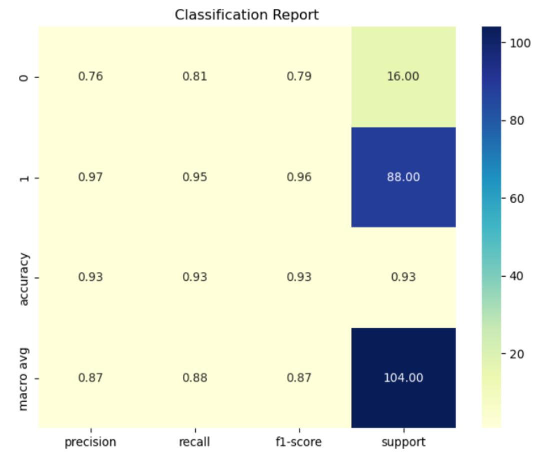


Fig. 5 Classification Report

Moreover, d_{50} is identified as the first splitting attribute, as shown in Figure 6, emphasizing d_{50} (average grain size) as the most important feature in the dataset. As mentioned, d_{50} had the most significant impact on the liquefaction prediction in the case of the Random Forest algorithm, with the groundwater level ranking second in feature importance. Lastly, fines content, natural moisture content, plasticity index, and standard penetration test value follow consecutively. The relative importance of each feature for predicting soil liquefaction susceptibility is given in Figure 6.

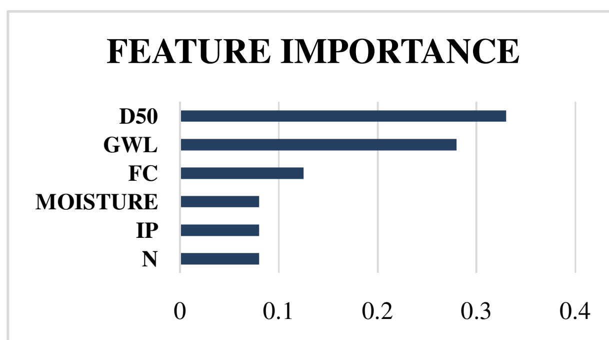


Fig. 6 Feature Importance

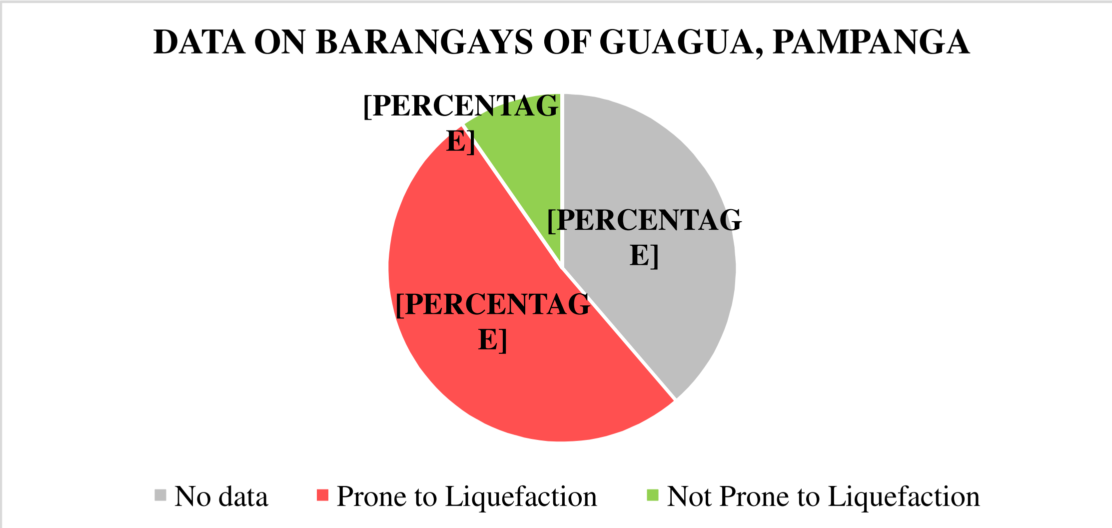


Fig. 7 Data on Barangays of Guagua, Pampanga

As for the second objective, this study aimed to produce a soil liquefaction susceptibility map of Guagua, Pampanga given the data collected from various stakeholders. The researchers were able to collect 26 SPT data conducted in Guagua, Pampanga that represents 19 barangays. After inputting the datasets into the working web application, it was found that 16 out of 31 barangays in Guagua, Pampanga were susceptible to soil liquefaction. In addition, 3 out of the 31 barangays were found safe against soil liquefaction. Lastly, there were no SPT data collected in the remaining 12 barangays of Guagua, Pampanga. The summary of data on soil liquefaction of barangays in Guagua, Pampanga is shown in Figure 7. The results of the second objective can also be summarized in Table 4. Even though 413 datasets where fed to the machine learning model, only 74 were utilized to map out the soil liquefaction susceptibilty of Guagua, Pampanga.

TABLE IV
SUSCEPTIBILITY TO SOIL LIQUEFACTION OF BARANGAYS IN GUAGUA, PAMPANGA

Barangay	**Susceptibility to soil liquefaction**	**Number of datasets**	**Datasets in percent**
Ascomo	Not susceptible	3	4.05%
Bancal	Not susceptible	3	4.05%
Jose Abad Santos	Unidentified	0	0.00%
Lambac	Susceptible	2	2.70%
Magsaysay	Susceptible	3	4.05%
Maquiapo	Susceptible	3	4.05%
Natividad	Susceptible	9	12.16%
Plaza Burgos	Unidentified	0	0.00%
Pulungmasle	Susceptible	9	12.16%
Rizal	Susceptible	3	4.05%
San Antonio	Susceptible	3	4.05%
San Agustin	Susceptible	3	4.05%
San Isidro	Susceptible	5	6.76%
San Jose	Unidentified	0	0.00%
San Juan	Not susceptible	3	4.05%
San Juan Bautista	Unidentified	0	0.00%
San Juan Nepomuceno	Unidentified	0	0.00%
San Matias	Susceptible	2	2.70%
San Miguel	Susceptible	3	4.05%
San Nicolas 1st	Unidentified	0	0.00%
San Nicolas 2nd	Susceptible	6	8.11%

San Pablo	Unidentified	0	0.00%
San Pedro	Unidentified	0	0.00%
San Rafael	Susceptible	3	4.05%
San Roque	Unidentified	0	0.00%
San Vicente	Susceptible	5	6.76%
Santa Filomena	Unidentified	0	0.00%
Santa Ines	Susceptible	3	4.05%
Santa Ursula	Susceptible	3	4.05%
Santo Cristo	Unidentified	0	0.00%
Sto. Nino	Unidentified	0	0.00%

Notably, most of the the barangay has at least two datasets. Ascomo, Bancal, Magsaysay, Maquaipo, Rizal, San Anonio, San Agustin, San Juan, San Miguel, San Rafael, Santa Ines, and Santa Ursula were the barangays that comprise 4.05% each in generating the hazard map. On the on other hand, there were two barangays that have 12.16% datasets, Natividad and Pulungmasle. Some barangays had only 2.70% in terms of datasets used in genrating the hazard map which are San Matias and Lambac. Also, barangays like San Isidro and San Vicente had datasets that amounted to 6.76 in terms of percent. Lastly, the lone barangay that had 8.11% is the barangay of San Nicoles 2nd. Unfortunaly, due to lack data, there were 11 barangays that had 0%, hence, the barangays that have an unidentified soil liquefaction susceptibility.

Finally, the last objective of this study was to develop a working web application that prints and determines the soil liquefaction potential of a certain area. In this section, the different features of the web application developed by the researchers with the help of a web developer are discussed in the succeeding paragraphs. The features include the Login Page, Navigation Bar, Dashboard Page, Administrator Page, Barangay Page, About Page, and Landing Page. However, the only feature that is accessible to the public is the landing page. On the other hand, all features were made to be accessible only by the administrators to maintain the security and the legitimacy of the hazard map.

B. Discussion

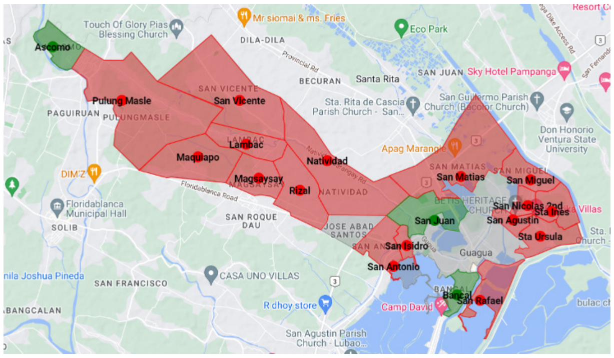
There is no existing web application to predict the susceptibility to liquefaction of a certain area up to this day. The researchers developed a web application that uses a machine learning algorithm to predict soil liquefaction in a barangay, based on crucial data. This web application accepts 6 input variables to determine the susceptibility to liquefaction of a barangay. The input variables include standard penetration test value (N-Value), natural moisture content (ω), plasticity index (PI), fines content (FC), average grain size (D50), and groundwater level (GWL). After inputting these variables, the model will determine the susceptibility by displaying “Yes” or “No”. Finally, once the name of the barangay and its liquefaction susceptibility are saved, this web app can print the hazard map of Guagua, Pampanga.

The model used in the web application achieved a high accuracy level in predicting liquefaction susceptibility. This is after subjecting the model to rigorous training using the Random Forest algorithm. Based on the classification report, the model attained a relatively high score for different performance indicators on accurately predicting the liquefaction susceptibility of a certain barangay. According to the results presented, it can be implied that the model is fitting to be used in assessing the liquefaction potential of a barangay. This indicates that the model can forecast soil liquefaction susceptibility in an accurate manner. Also, this implies that there is only a small room for error when making predictions with those datasets that are susceptible to soil liquefaction. However, it is advised to still be cautious at all times and not to completely rely on the model’s ability to make accurate predictions. Furthermore, the model scored 80% on average for datasets without liquefaction which is on the low side compared to those with liquefaction. This simply means that the model has trouble recognizing datasets that are not susceptible to liquefaction and displays them as susceptible. Even though this causes others to be more cautious, this reason does not outweigh the fact that the model makes inaccurate predictions on the no liquefaction side. However, this issue can be solved by adding more no liquefaction datasets.

The model has been developed using the 413 provided datasets, and parameters were derived through cross-validation. Ten of the best combinations derived from the cross-validation of the dataset are given in Table 4. The parameters obtained from the Random Forest algorithm will be applied to the complete dataset and incorporated into the web application developed in the study. Moreover, the model obtained the best accuracy rate shown in Figure 6. For precision, 76% of the no liquefaction datasets were accurately predicted and 97% of datasets with liquefaction were accurately predicted. As for recall, out of all the datasets, only 81% were predicted by the model for no liquefaction and 95% for susceptible to liquefaction. Lastly, for f1-score, the model scored 79% for no liquefaction and 96% for susceptible to liquefaction. Overall, the result obtained when the datasets were trained with the algorithm reached an accuracy of 93%. All of these were supported by 104 training datasets. Performance evaluations were made according to the average test accuracy score. According to the results given, it can be seen that the predictive model gave a high accuracy rate with 93% in the training set and approximately 96% in the dataset. This suggests that the model is well-suited for predicting soil liquefaction susceptibility.

Based on a research that utilized random forest in evaluating soil liquefaction, but using different parameters—specifically, the shear wave velocity—the random forest model trained achieved above 90% accuracy in predicting datasets with actual soil liquefaction. Furthermore, their model also scored less than 80% with those datasets with non-liquefied soils. This further proves that the accuracy of the model used in generating the soil liquefaction hazard map for the Municipality of Guagua, Pampanga was appropriate and adequate On the other hand, another study that used random forest algorithm and variables such as friction ratio, peak horizontal acceleration, vertical effective stress, cone penetration resistance, and frictional resistance, the model was found to be accurate in predicting both liquefied and non-liquefied soil cases and achieved an accuracy rate of 98.4% [35]. This means that the model trained is less accurate compared to this study that utilized other input variables. However, this can also be interpreted that the model has rooms for improvements. As long as there is more time and data, generating a more accurate soil liquefaction hazard map will not be impossible. It is also worth to mention that since this study utilized the random forest algorithm, meaning, the model's decisions were based on numerous decision trees. Therefore, no flowchart-based criteria will be produced.

Guagua, Pampanga has soft soil due to its geographical location and the lahar deposited by Mt. Pinatubo in 1991. As such, the researchers developed a web application that can print the hazard map of Guagua, Pampanga concerning soil liquefaction. In Figure 19, it can be seen that most areas of Guagua, Pampanga are highlighted with the color red. This means that most of the barangays in Guagua are susceptible to liquefaction. Based on the collected data, only Ascomo, San Juan, and San Isidro are the barangays that are not susceptible to liquefaction. However, since the model has an issue in predicting the liquefaction of the barangay that has no liquefaction, San Isidro was reflected as red. Out of the 26 datasets for 19 barangays, 25 were predicted accurately, this gives the researchers 96.15% accuracy, almost the same as the accuracy score of the predictive model.



Legend:

No Data

Prone to Liquefaction

Not Prone to Liquefaction

Fig. 8 The Soil Liquefaction Hazard Map of Guagua, Pampanga

Furthermore, according to the HazardHunterPH, a joint project with several government institutions in the lead DOST, the municipality of Guagua, Pampanga in general is highly susceptible to soil liquefaction which is also the general conclusion of this study[36]. The HazardHunterPH also provides a soil liquefaction hazard map, however, this web application is not dynamic. The hazard map generated from the HazardHunterPH was printed in 2010. Unlike the web application of this study, the web application is synchronized with the administrators input which ensures an up-to-date hazard map. In addition, GeoAnalyticsPH is an innovative web application that allows users to generate informative maps and analytics using data from the GeoRiskPH database. This includes information about hazards, exposure, and location. Users can gain a better understanding of data by visualising it with maps, charts, and graphs, allowing them to plan ahead of time for natural disasters [37]. In the same way, the web application of this study offers the same features, however due to limited data the researchers were not able to show a detailed information regarding liquefaction.

IV. CONCLUSIONS

In this study, the soil liquefaction hazard map of Guagua, Pampanga was created through a web application that was incorporated with a predictive model trained using random forest—a kind of machine learning algorithm. As per the study's objectives, the predictive model was developed using 413 collected soil liquefaction datasets from various testing centers and DPWH districts in Pampanga. The model was trained using random forest algorithm and achieved an accuracy of 93%. Consequently, the model was incorporated into a web application that was developed using Joblib and SQLAlchemy, both of which are Python-based libraries. This web application allows users to easily input their geotechnical data and receive an accurate estimate of the susceptibility of soil liquefaction, thus providing the soil liquefaction hazard map of Guagua, Pampanga, as shown in Figure 19. From the results, it can be concluded that machine learning algorithms can be utilized to assess soil liquefaction potential, provided that there are available datasets for the models to train with. In addition, the result of this contributes to other geotechnical and earthquake engineering research studies.

The hazard map, generated from the results of this study, will serve as a tool to facilitate better planning and minimize the risk of soil liquefaction in Guagua, Pampanga. With the current data that the researchers collected, 16 out of 19 barangays were found to be susceptible to soil liquefaction. This represents the 84.21% of the collected data of barangays in Guagua, Pampanga and 51.61% of all of the barangays in Guagua, Pampanga. On the other hand, 3 of the 19 collected data of barangays in Guagua, Pampanga were found to be safe against soil liquefaction. This represents 15.79% and 9.68% of collected barangays and all barangays, respectively, in Guagua, Pamanga. However, the liquefaction susceptibility of 12 barangays in Guagua, Pampanga remain unidentified which represents 38.71% of the barangays in Guagua, Pampanga. With that, it can be concluded that the Municipality of Guagua, Pampanga has high probability of experiencing soil liquefaction. Lastly, the hazard map produced by this study should not be utilized for site specific investigation but can be used in risk management and emergency planning and investigation.

ACKNOWLEDGEMENTS

The researchers would like to express their gratitude to the people who have participated in this study and to the following individuals who has helped them in completing the study with sufficient data and factual information.To the Almighty God, for bestowing guidance and grace in accomplishing the research study.Their parents, who gave their endless love, prayers, support, and sacrifices. None of this would have been possible without their love and encouragement.Their research adviser, Ma. Angelu S. Castro, RCE, RMP, for leading her students towards the significance of measurement accuracy in data presentation. She has shared her great knowledge in the field of geothecnical engineering. Their research coordinator, Charles G. Lim, ECE, ECT, for sharing invaluable guidance, support, advice, comments, suggestions, and provisions that helped the completion and success of this study. The DHVSU Civil Engineering Faculty, for offering their constant support, advice, and patience in answering the queries of the researchers in their study. The researchers would also want to acknowledge the DPWH 2nd District Engineering Office, the DPWH Regional Office, and the Unified Geotest Laboratory in giving us the necessary data in completing the study.And lastly, to DOST, for offering support financially to the researchers.

REFERENCES

[1] O. Selcukhan and A. Ekinci, 'Assessment of Liquefaction Hazard and Mapping Based on Standard Penetration Tests in the Long Beach and Tuzla Regions of Cyprus', *Infrastructures*, vol. 8, no. 6, May 2023, doi: 10.3390/infrastructures8060099.

[2] M. Dhahir, W. Nadir, and M. Rasool, 'Influence of Soil Liquefaction on the Structural Performance of Bridges During Earthquakes: Showa Bridge as A Case Study', *Sci. Publ. Corp.*, vol. 7, no. 20, Nov. 2018, doi: 10.14419/ijet.v7i4.20.25916.

[3] J. Galupino and J. Dungca, 'Estimating Liquefaction Susceptibility Using Machine Learning Algorithms with a Case of Metro Manila, Philippines', *Appl. Sci.*, vol. 13, no. 11, May 2023, doi: 10.3390/app13116549.

[4] I. Jaime, J. Panhugban, J. B. Tabora, and A. Pineda, 'Geo-hazard Mapping of the Province of Pampanga: A Reference for Disaster Preparedness Panorama', issuu. Accessed: Nov. 26, 2023. [Online]. Available: https://amp.issuu.com/uasbpa/docs/sbpa_journal/s/12385116

[5] L. Z. Mase, M. Farid, N. Sugianto, and S. Agustina, 'The Implementation of Ground Response Analysis to Quantify Liquefaction Potential Index (LPI) in Bengkulu City, Indonesia', *J. Civ. Eng. Forum*, vol. 6, no. 3, Sep. 2020, doi: 10.22146/jcef.57466.

[6] W. Dwiyantoro, T. F. Fathani, and A. D. Adi, 'Influence of Groundwater Table Fluctuation on Liquefaction Potential Analysis Using Cyclic Stress Approach', *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1184, no. 1, p. 012006, May 2023, doi: 10.1088/1755-1315/1184/1/012006.

[7] R. Bade, S. Bhoyar, and P. Sonekar, 'Earthquake Induced Liquefaction around the World', vol. 9, no. 5, May 2021, [Online]. Available: https://ijcrt.org/papers/IJCRT21A6004.pdf?fbclid=IwAR24X0x4B-_DjVy0T1ZYPZnJLfcWeqwGJYFTmHZQTfiCIPwQGFgJoGTwmjw

[8] Earthquake Hazard Program, 'San Francisco Bay Area Liquefaction Hazard Maps', United States Geological Survey. Accessed: Nov. 26, 2023. [Online]. Available: https://www.usgs.gov/programs/earthquake-hazards/science/san-francisco-bay-area-liquefaction-hazard-maps#overview

[9] F. Abdurachman, A. Dean, and R. Paddock, 'A Tsunami Didn't Destroy These 1,747 Homes. It Was the Ground Itself, Flowing.', *New York Times*, Indonesia, Oct. 18, 2023. Accessed: Nov. 10, 2023. [Online]. Available:

https://www.nytimes.com/2018/10/03/world/asia/indonesia-earthquake-tsunami-liquefaction.html

[10] J. Daniell, A. Schaefer, and F. Wenzel, 'Losses Associated with Secondary Effects in Earthquakes', *Front. Built Environ.*, vol. 3, Jun. 2017, doi: 10.3389/fbuil.2017.00030.

[11] H. Kojima and K. Tokimatsu, 'Liquefaction-induced damage and geological and geophysical conditions during the 1990 Luzon earthquake'. Akio Abe, Tokyo Soil Research Co., Ltd., Japan, 1992. [Online]. Available: https://www.iitk.ac.in/nicee/wcee/article/10_vol1_135.pdf?fbclid=IwAR2CKMVnhbv5HNghRxS-C3P-mX0K2tHLYR_7uPpLzH4zw6_FbO8RypCg9VA

[12] G. Lacorte and O. Dinoy, 'Liquefaction sank Davao Sur building', *Inquirer News*, Davao del Sur, Dec. 21, 2019. Accessed: Nov. 26, 2023. [Online]. Available: https://newsinfo.inquirer.net/1203913/liquefaction-sank-davao-sur-building?fbclid=IwAR0K7m0Pf3jElhvWCJhpxPHH3dhJ-KQP3Go5QZ2kHXzanFO2aTfkm6O_9xA

[13] PHIVOLCS, 'PRIMER ON THE 22 APRIL 2019 MAGNITUDE 6.1 CENTRAL LUZON EARTHQUAKE', PHIVOLCS. Accessed: Nov. 26, 2023. [Online]. Available: https://www.phivolcs.dost.gov.ph/index.php/news/8233

[14] M. Lopez, 'Why Pampanga was hit hardest by earthquake when epicenter was in Zambales', *CNN Philippines*, Zambales, Apr. 23, 2019. Accessed: Nov. 26, 2023. [Online]. Available: https://www.cnnphilippines.com/news/2019/4/23/Pampanga-earthquake-effects.html

[15] iMedia, 'Imitation versions of landmarks from other countries in the world in the Philippines, the last one is very happy', iMedia. Accessed: Nov. 26, 2023. [Online]. Available: https://min.news/en/travel/b76067f33a4f15f36dc0d692a7f141ff.html?fbclid=IwAR1Btt-gHrttP4D0AQwfOfT3KHvWbtNB5tSn9yszP2Xw-SHhIjM8atlbXp0

[16] V. Ferreras, 'DPWH builds action center in Pampanga should "Big One" hit NCR', *CNN Philippines*, NCR, Feb. 14, 2023. Accessed: Dec. 05, 2023. [Online]. Available: https://www.cnnphilippines.com/news/2023/2/14/DPWH-action-center-Pampanga-Big-One-NCR-.html#google_vignette

[17] G. Bongon, 'The Big One: Unraveling the Impending Earthquake in the Philippines', Linked in. Accessed: Dec. 05, 2023. [Online]. Available: https://www.linkedin.com/pulse/big-one-unraveling-impending-earthquake-philippines-gerry-bongon

[18] J. Aguila, 'How strong an earthquake will the "Big One" be? You need to sit down for the answer', *ABS-CBN News*, May 02, 2019. Accessed: Dec. 05, 2023. [Online]. Available: https://news.abs-cbn.com/ancx/culture/spotlight/05/02/19/should-you-be-afraid-of-the-big-one

[19] PHIVOLCS, 'FaultFinder', FaultFinder. Accessed: Nov. 26, 2023. [Online]. Available: http://faultfinder.phivolcs.dost.gov.ph/

[20] Place and See, 'Guagua', placeandsee. Accessed: Dec. 02, 2023. [Online]. Available: https://placeandsee.com/wiki/guagua?fbclid=IwAR1FvZQMldbUDyQg8JjqPdbb7h58gFG2BjCSvW7-pNdGi2ipafjrVyYxhxA

[21] PhilAtlas, 'Guagua, Pampanga Profile', PhilAtlas. Accessed: Dec. 02, 2023. [Online]. Available: https://www.philatlas.com/luzon/r03/pampanga/guagua.html?fbclid=IwAR0ItDVdqKrCSVsyH8gErh-0d72q2KDjpcvz7xOpKn2OFSrTwD397CNPOcM

[22] Department of Trade and Industires, 'Pampanga Profile - Cities and Municipalities Competitive Index', Cities and Municipalities. Accessed: Jan. 05, 2024. [Online]. Available: https://cmci.dti.gov.ph/prov-profile.php?prov=Pampanga&fbclid=IwAR14esB1cmVu0qJlBZTc3Mf2ba5ljGyNCGtGheKiBJwx7wOnr3jxeNwCRZ0

[23] C. M. Bishop, *Pattern recognition and machine learning*. in Information science and

statistics. New York: Springer, 2006. [Online]. Available: http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf

[24] L. Breiman, 'Random forests', *Kluwer Acad. Publ.*, vol. 45, pp. 5–32, 2001.

[25] V. R. Kohestani, M. Hassanlourad, and A. Ardakani, 'Evaluation of liquefaction potential based on CPT data using random forest', *Nat. Hazards*, vol. 79, no. 2, pp. 1079–1089, Nov. 2015, doi: 10.1007/s11069-015-1893-5.

[26] A. Biswal, 'Data Analytics With Python: Use Case Demo', Simplilearn.com. Accessed: Mar. 30, 2024. [Online]. Available: https://www.simplilearn.com/tutorials/data-analytics-tutorial/data-analytics-with-python

[27] J. Terra, 'Why Python Is Essential for Data Analysis and Data Science', Simplilearn.com. Accessed: Mar. 30, 2024. [Online]. Available: https://www.simplilearn.com/why-python-is-essential-for-data-analysis-article

[28] P. Kaur, 'How to Build an API With Python Flask', Moesif Blog. Accessed: Mar. 30, 2024. [Online]. Available: https://www.moesif.com/blog/technical/api-development/Building-RESTful-API-with-Flask/

[29] A. Zola, 'What is a Bootstrap and how does it work?', Tech Target. Accessed: Mar. 30, 2024. [Online]. Available: https://www.techtarget.com/whatis/definition/bootstrap

[30] A. Tharwat, 'Classification assessment methods', vol. 17, no. 1, Jul. 2020, doi: 10.1016.

[31] Evidently AI, 'How to interpret a confusion matrix for a machine learning model', Evidently AI. Accessed: Apr. 21, 2024. [Online]. Available: https://www.evidentlyai.com/classification-metrics/confusion-matrix

[32] Scikit-Learn, 'F1 score', scikit-learn. Accessed: Apr. 21, 2024. [Online]. Available: https://scikit-learn/stable/modules/generated/sklearn.metrics.f1_score.html

[33] Pragmatic Institute, 'Overcoming the 80/20 Rule in Data Science | Pragmatic Institute', Pragmatic Institute - Resources. Accessed: Apr. 01, 2024. [Online]. Available: https://www.pragmaticinstitute.com/resources/articles/data/overcoming-the-80-20-rule-in-data-science/

[34] E. Zvornicanin, 'What Is Feature Importance in Machine Learning?', Baeldung. Accessed: Apr. 21, 2024. [Online]. Available: https://www.baeldung.com/cs/ml-feature-importance

[35] R. B. Bhardwaj and S. R. Chaurasia, 'Use of ANN, C4.5 and Random Forest Algorithm in the Evaluation of Seismic Soil Liquefaction', *J. Soft Comput. Civ. Eng.*, vol. 6, no. 2, pp. 92–106, Apr. 2022, doi: 10.22115/scce.2022.314762.1380.

[36] PHIVOLCS, 'HazardHunterPH - Hazard assessment at your fingertips'. Accessed: May 22, 2024. [Online]. Available: https://hazardhunter.georisk.gov.ph/

[37] D. Reyes, 'FREE TOOL: Using "Map Data" to Measure Hazards and Risks Wherever You Are in the Philippines', Engineer Dee's Blog. Accessed: May 30, 2024. [Online]. Available: https://engineerdee.com/georisk-philippines/