International Journal of GEOMATE, June 2019, Vol.16, Issue 58, pp.48 - 54
Geotec., Const. Mat. & Env., DOI: https://doi.org/10.21660/2019.58.8129
ISSN: 2186-2982 (Print), 2186-2990 (Online), Japan

QUEZON CITY SOIL PROFILE REFERENCE

Joenel G. Galupino[1] and Jonathan R. Dungca[1]

[1]Civil Engineering Department, De La Salle University, Manila, Philippines

*Corresponding Author, Received: 23 Oct. 2018, Revised: 04 Dec. 2018, Accepted: 20 Dec. 2018

ABSTRACT: The City of Quezon City is one of the highly urbanized cities and one of the fastest growing metropolitan areas in the Philippines, many local and foreign investors are discovering it as a cost-effective business location; many infrastructures were built to serve these growing business hub. Every infrastructure project constructed rests on the ground, without knowing the soil interaction underground, safety is at risk. Thus, this study aims to generate the soil profile of Quezon City using machine learning, specifically, k-Nearest Neighbor (k-NN) algorithm; k-Nearest Neighbor (k-NN) measured the similarity of soil types in terms of distance. The soil profile generated by the model was delineated using computer-aided design (CAD); it was discovered that the underground of the Quezon City is usually dominated by tuff. The generated soil profile will not only serve engineers to decide what type of foundations to be used for a particular site but will also be used for Disaster Risk Reduction (DRR) planning to mitigate ground related disasters; government zoning and policymakers for land use purposes; for real estate industry as their initial reference before investing. The nearest neighbor algorithm model used in the generation of the soil profiles was cross-validated to ensure the predictions are adequate.

Keywords: Machine learning, Soil profile, Nearest neighbor, Philippines, Borehole

1. INTRODUCTION

Every infrastructure project constructed need foundations to stand and these foundations are situated underground. Engineers need to study the soil interaction to economically and adequately design these foundations but without conducting on-site tests, Engineers cannot clearly describe the said interactions underground. In order to reduce the cost of the project, some Engineers rely on previous explorations nearest to project site to approximate the soil properties since these tests are very expensive [1] and some of the soils were stabilized using waste materials [3-10]. Usually, these soils are heterogeneous and it is characterized through a profile by dividing it into horizons based on properties observed in the field [2].

Quezon City is one of the highly urbanized cities of Metro Manila, Philippines, and it is also one of the fastest growing metropolitan areas in the Philippines. Quezon City is a growing enterprise hub, with 58,000 registered business mostly in line with retail trade, restaurants, contractors of goods and services, manufacturers and amusement places. The yearly average of business applicants totaled 11,000 with about 43 establishments registered daily. Many local and foreign investors are discovering Quezon City as a cost-effective business location, shopping malls and a huge Information Technology (IT) parks are being constructed [11], therefore, structures are being constructed to house these developments, and as mentioned, soil explorations are pre-requisite to these projects.

To eliminate the previous information requirement about the interactions among inputs, parameters, and outputs in soil explorations, different soft computing techniques have effectively been applied. To tackle the limitations of numerical and empirical models, artificial neural network (ANN), fuzzy inference systems (FIS), adaptive-network-based fuzzy inference system (ANFIS), Bayesian network (BN) and genetic programming (GP) are the most common methods being applied. In this study, the k-nearest neighbor (k-NN) algorithm was used because it is mainly employed for measuring the similarity of a set of the object based on some measures of distance and is one the oldest pattern classifier methods with no required pre-processing [12].

Furthermore, in the dawn of new technology, data are being captured, stored, manipulated, analyzed, managed, and presented. One of the systems used is the Geographical Information System (GIS), it is a system that integrates, stores, edits, analyzes, shares, and displays geographic information which can be used in the different technologies, processes, and methods [13]. In the soil profile research, Geographical Information System is a great tool, GIS solves the problematic field delineation of soil horizons.

Thus, this study aims to generate the soil profile of Quezon City using a k-NN algorithm that will not only serve as a reference for Engineers but will also serve as a guide for policymakers in Quezon City, Philippines.

2. METHODOLOGY

Soil borehole logs located in Quezon City was collected and was plotted in a map, shown on Fig. 1. A density of one borehole log per square kilometer was used to describe the soil profile. The distribution was visually inspected and the areas that needed more data were determined. Borehole logs that seemed erroneous were also removed and disregarded.

Fig. 1. Borehole Locations of Quezon City

The elevations of each borehole were also plotted because these points are the references for the soil profile characterization, a 3D Elevation Map of Quezon City is shown in Fig. 2. Furthermore, a 16 by 16 grid was used in the study to represent each borehole, shown on Fig. 3.

Fig. 2. 3D Elevation Map of Quezon City

The soils are grouped into classes, which has similar physical properties and general characteristics in terms of behaviors. The grouping system is usually related to its physical properties inherent in the soil and not for a particular use. With only the soil type available it is not sufficient for design purposes but it will give the engineer an indication of the behavior of soil when used as a component in construction [14]. The soil types are usually grouped by the following [15], shown in Table 1:

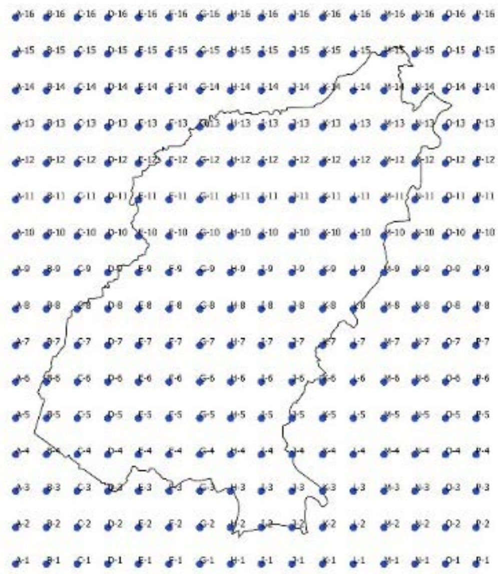

Fig. 3. A grid of Quezon City

A k-Nearest Neighbor (k-NN) algorithm model was used because is one of the simplest classification algorithms and it is one of the most used learning algorithms. It created a model that used the borehole log database in which the data points were separated into several groups to predict the classification of a new sample point. Once the grids have been deployed, the results of the k-NN model will be used in the creation of soil profile using GIS and/or CAD. Each k-NN model consists of a data case having a set of independent variables labeled by a set of dependent outcomes, the research k-NN model classification of is shown in Fig. 4. The shapes signify a particular soil type, shown in Table 1. The independent and dependent variables can be either continuous or categorical. In the study, the dependent and independent variables are shown in Table 2.

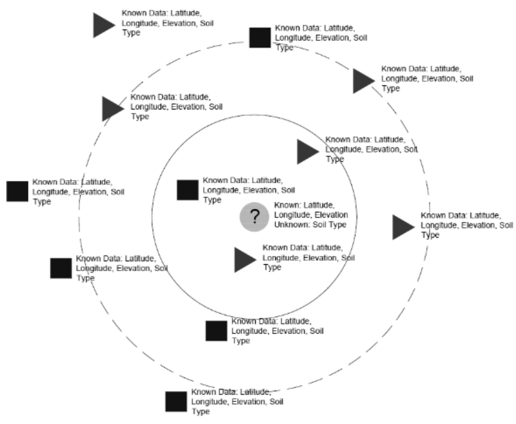

Fig. 4. Research k-NN model classification

Table 1. Soil Classification [15]

Soil Type	Description
Sand	Easy to compact and has little effect by moisture. It has the same properties as gravel, the only difference and the division is the No.4 sieve. Rounded to angular, bulky, hard, rock particle, passing No. 4 sieve (4-76 mm) retained on No. 200 sieve (0-74 mm).
Clay	Cohesive soil increases with a decrease in moisture. It is usually subjected to expansion and shrinkage with changes in moisture. Its permeability is very low and impossible to drain by regular means. Particles smaller than No. 200 sieve (0-74 mm) identified by behavior; that is, it can be made to exhibit plastic properties within a certain range of moisture and exhibits considerable strength when air dried.
Silt	Has a tendency to become quick when saturated and is inherently unstable. It is easily erodible and subject to piping and boiling. Particles smaller than No. 200 sieve (0-74 mm) identified by behavior; that is, slightly or non-plastic regardless of moisture and exhibits little or no strength when air dried.
Gravel	Like sand, also easy to compact and has little effect by moisture Gravels are generally more pervious and resistant to erosion and piping compared to sands. Rounded to angular bulky, hard, rock particle, passing 3-in. sieve (76-2 mm) retained on No. 4 sieve, (4-76 mm).
Tuff	Construction material which is generally a limestone precipitated from groundwater. In soil exploration reports, it is a rock mass and its core recovery are usually measured in the rock-quality designation (RQD). Usually, the foot of foundations rest on tuffs

An estimate of k can be determined using cross-validation. Cross-validation is a well-established technique that can be used to obtain estimates of model parameters that are unknown [16].

Table 2. Dependent and Independent Variables

Independent Variable(s)	Dependent Variable(s)
1. Soil Type	1. Longitude 2. Latitude 3. Elevation

As mentioned, one can make a decision on the class of a sample (query) according to the calculated similarities with the k-NN. Usually, the dependency is computed using Euclidean distance which is defined in Eq. 2 [17]:

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \qquad (1)$$

Where: d(x,y) is the Euclidean distance between the samples x and y, which could have n dimensions in the feature space.

k-NN predictions are based on the intuitive assumption that objects close in distance are potentially similar, it makes good sense to discriminate between the K nearest neighbors when making predictions [16]. By introducing a set of weights W, the closest points among the k nearest neighbors have more influence in affecting the outcome of the query point, shown on Eq. 2:

$$W(x,p_i)\frac{e^{-D(x,p_i)}}{\sum_{i=1}^{K} e^{-D(x,p_i)}} \qquad (2)$$

Where: $D(x,p_i)$ is the distance between the query point x and the i^{th} case p_i of the example sample.

Once, the k-NN model has been established, the grid points were deployed per elevation, from -23m amsl to 88m amsl, shown in Fig. 5.

In a particular study [10] which utilized the same machine learning model, it was able to provide a *k*-nearest neighbor model that served as a reference to predict the compressive strength of concrete while incorporating waste ceramic tiles as a replacement to coarse aggregates while varying the amount of fly ash as a partial substitute to cement.

3. ANALYSIS AND DISCUSSION

Quezon City is a landlocked city bordered by Caloocan and Valenzuela City to the west and northwest and Manila to the southwest. San Juan and Mandaluyong to the south lie and Marikina and Pasig border the city to the southeast. San Jose del Monte in the province of Bulacan to the north and Rodriguez and San Mateo, both in the province of Rizal to the east.

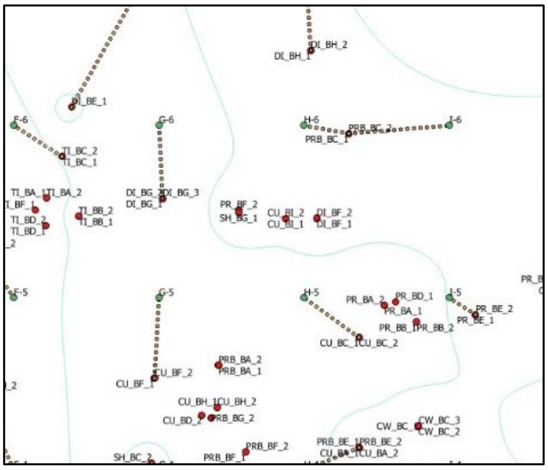

Fig. 5. Deployment of 16x16 grid to the k-NN model

A relatively high plateau at the northeast of the metropolis situated between the lowlands of Manila to the southwest and the Marikina River Valley to the east, known as Guadalupe Plateau, is where Quezon City lies.

Manila, where Quezon City is located, was submerged at one time in the geologic past. Intermittent volcanic activities followed and after which, volcanic materials were deposited [1]. Volcanic rocks, known as "Adobe", is the common rock in the underlying layers. It is locally known as the Guadalupe Formation, it is composed of Lower Alat Conglomerate Member and the Upper Diliman Tuff Member. The Diliman Tuff includes the tuff sequence in the Angat-Novaliches region and along Pasig River in the vicinity of Guadalupe, Makati and extending to some areas of Manila and most of Quezon City [1].

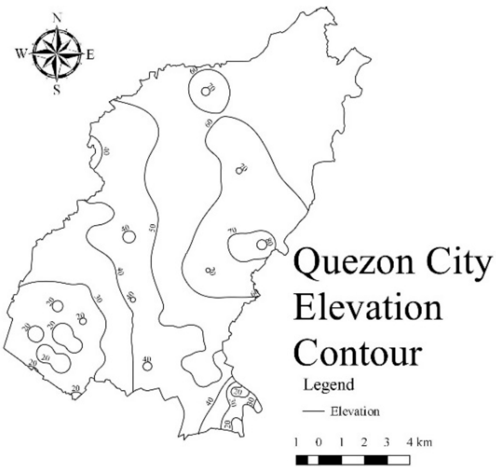

Fig. 6. Elevation Contour of Quezon City

The maximum ground elevation of Quezon City is 88 meters above mean sea level (msl) located at Diliman, Quezon City, and the lowest is 5 meters above mean sea level is at Sta. Mesa border, the Elevation Contour Map is shown in Fig. 6. The minimum level of soil profile elevation was influenced by a particular borehole located near Sta. Mesa which has a ground elevation of 7 meters above msl and its borehole was 30 meters deep, thus, in the study, a minimum of 23 meters (-23m amsl) below msl was used.

Quezon City's soil profile has prevalent tuff layers. Cities with rock formations beneath the surface, such as Quezon City have soils with high bearing capacities at shallow depths. It is recommended to place the foundations on these refusal levels since it is more than capable of carrying loads that are suited for shallow foundations [1].

3.1 Longitudinal

The soils are usually grouped into classes as mentioned, groups are based on similar physical properties and general characteristics in terms of behaviors.

Tuff is very prevalent in Quezon City, just several meters below the ground, it would require from Soil Penetration Test (SPT) to Rock Quality Designation (RQD) to accommodate the tuff lying beneath the clay and sand surface, a sample delineation is shown on Fig. 7. There are gravel and clay layers in between the tuff layers, it would not greatly affect the soil bearing capacity of the tuff layer since tuff can handle a large amount of load.

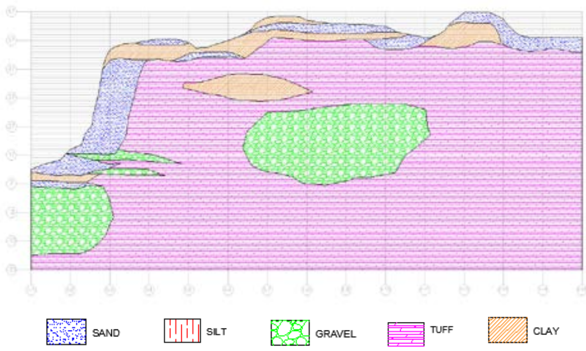

Fig. 7. Soil Profile in the Longitudinal Grid I

Along the longitudinal grid, it was mentioned that there are prevalent tuff layers. Grids C, F, I and L are 78%, 97%, 86% and 29% composed of Tuff, respectively, a chart is shown in Fig. 8. Also in Grids L-O, gravelly layers are prevalent which composes 60% and 84%, respectively. It should be expected that along the Longitudinal Grids, shallow foundations may be recommended. An alternating layer of sand and clay is also prevalent in the shallow layers is due to sediment deposits that are left by the rivers and creeks over time [1].

3.2 Traverse

Tuff layer also dominates the underground surface. There are clay, silt and sand layers in the top but several meters below the ground, refusal levels may be expected. Shallow foundations may also be recommended along with the Traverse grid. A sample soil profile in the traverse grid is shown in Fig. 9.

Along the traverse grid, a layer of tuff was dominant. Grids 3, 6, 9, 12 and 15 are 89%, 82%, 63%, 47% and 51% Tuff, a chart is shown on Fig. 10. Thus, with the prevalence of Tuff, it should be expected that along the Traverse Grids, shallow foundations may also be recommended. Same with the longitudinal grids, an alternating layer of sand and clay is also prevalent in the shallow layers is due to sediment deposits that are left by the rivers and creeks over time [1].

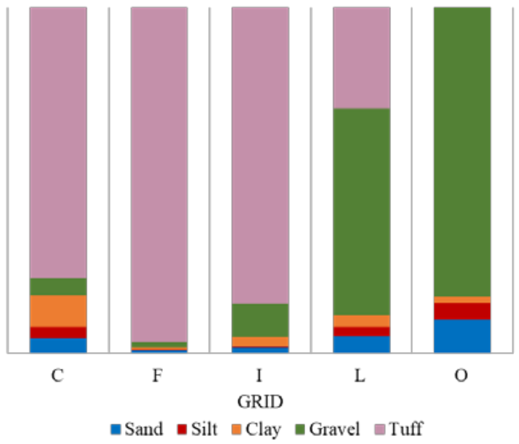

Fig. 8. Soil Types in the Longitudinal Grid

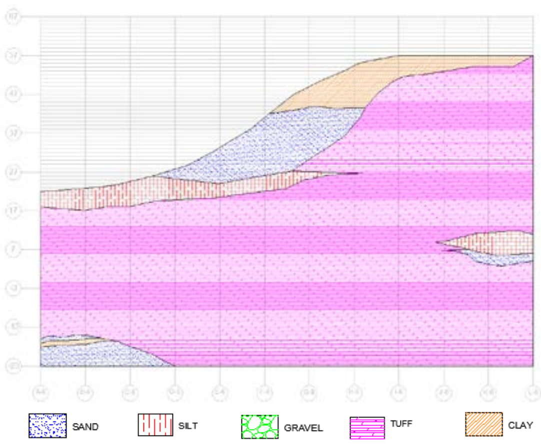

Fig. 9. Soil Profile in the Traverse Grid 6

3.3 k-Nearest Neighbor Analysis

k-Nearest Neighbor (k-NN) algorithm is one of the simplest classification algorithms and it is one of the most used learning algorithms. Each k-NN model consists of a data case having a set of independent variables labeled by a set of dependent outcomes. The specifications of the KNN analysis, including the list of variables selected for the analysis and the size of Example, Test, and Overall samples are shown in Table 3.

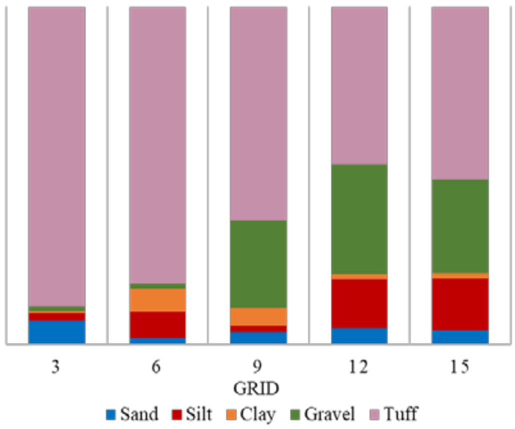

Fig. 10. Soil Types in the Traverse Grid

Table 3. Specifications of the k-NN Analysis

Specifications
Dataset Quezon City KNN.sta: Dependent: Soil Type Independents: Latitude, Longitude, Elevation Sample size = 2880 (Examples), 961 (Test), 3841 (Overall) KNN results: Number of nearest neighbors =49 Distance measure: Euclidean Averaging: uniform Cross-validation accuracy (%) = 75%

The independent variables are standardized to result in typical case values which fall into the same range. This will prevent independent variables with typically large values from biasing predictions.

With the prevalence of Tuff layers in Quezon City, in the KNN analysis, it is also expected that Tuff will have a higher confidence rate, followed by silt and sand which dominates the top layers of the soil profile. An Elevation vs. Confidence per Soil type is shown in Fig. 11:

3.4 Validation

The optimum K has the lowest test error rate. In the k-NN analysis, the model has been forced to fit the test set in the best possible manner. The training set is randomly divided into k groups, or folds, of approximately equal size. The first fold is treated as a validation set, and repeated k times; each time, a different group of observations is treated as a validation set. This process results in k estimates of the test error which are then averaged out, the cross-validation accuracy is shown in Fig. 12.

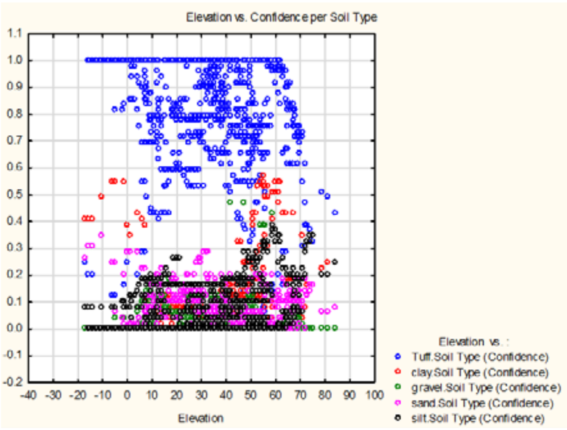

Fig. 11. Elevation vs. Confidence per Soil type

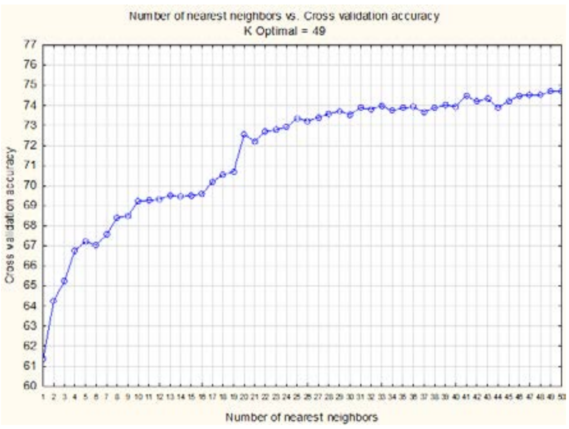

Fig. 12. Cross-Validation Accuracy

A cross-validation accuracy of 75% and a k-optimal of 49 was garnered which indicate that there is a strong relationship between the k-Nearest Neighbor (k-NN) model and the observed data.

4. CONCLUSION

This study was able to generate the soil profile of Quezon City using k-NN algorithm that will not only serve as a reference for Engineers but will also serve as a guide for policy makers in Quezon City, Philippines.

The soils are grouped into classes based on similar physical properties and general characteristics in terms of behaviors. Tuff is very prevalent in Quezon City, just several meters below the ground, it would require from Soil Penetration Test (SPT) to Rock Quality Designation (RQD) to accommodate the tuff lying beneath the clay and sand surface. It should be expected that shallow foundations may be recommended.

5. REFERENCES

[1] Dungca, J. R., Concepcion, I., Limyuen, M., See, T., & Vicencio, M. (2017). Soil bearing capacity reference for Mtero Manila, Philippines. International Journal of GEOMATE, 12(32), 5-11.

[2] Grauer-Gray, J., & Hartemink, A. (2018). Raster sampling of soil profiles. Geoderma, 99-108.

[3] Dungca, J.R., Jao, J.A.L. (2017). Strength and permeability characteristics of road base materials blended with fly ash and bottom ash. International Journal of GEOMATE, 12 (31), pp. 9-15.

[4] Dungca, J., Galupino, J., Sy, C., Chiu, A.S.F. (2018). Linear optimization of soil mixes in the design of vertical cut-off walls. International Journal of GEOMATE, 14 (44), pp. 159-165.

[5] Dungca, J.R., Galupino, J.G., Alday, J.C., Barretto, M.A.F., Bauzon, M.K.G., Tolentino, A.N. (2018). Hydraulic conductivity characteristics of road base materials blended with fly ash and bottom ash. International Journal of GEOMATE, 14 (44), pp. 121-127.

[6] Dungca, J.R., Galupino, J.G. (2017). Artificial neural network permeability modeling of soil blended with fly ash. International Journal of GEOMATE, 12 (31), pp. 76-83.

[7] Dungca, J.R., Galupino, J.G. (2016). Modeling of permeability characteristics of the soil-fly ash-bentonite cut-off wall using response surface method. International Journal of GEOMATE, 10 (4), pp. 2018-2024.

[8] Galupino, J.G., Dungca, J.R. (2015). Permeability characteristics of soil-fly ash mix. ARPN Journal of Engineering and Applied Sciences, 10 (15), pp. 6440-6447.

[9] Dungca, J.R., Codilla, E.E.T., II. (2018). Fly-ash-based geopolymer as a stabilizer for silty sand embankment materials. International Journal of GEOMATE, 14 (46), pp. 143-149.

[10] Elevado, K. T., Galupino, J. G., & Gallardo, R. S. (2018). Compressive strength modeling of concrete mixed with fly ash and waste ceramics using K-nearest neighbor algorithm. International Journal of GEOMATE, 15(48), 169-174.

[11] Casanova-Dorotan, F. (2010). Informal Economy Budget Analysis in the Philippines and Quezon CIty. Cambridge, MA: Women in Informal Employment: Globalizing and Organizing (WIEGO).

[12] Nikoo, M., Kerachian, R., & Alizadeh, M. (2017). A fuzzy KNN-based model for significant wave height prediction in large lakes. Oceanologia.

[13] Galupino, J., Garciano, E., Dungca, J., & Paringit, M. (2017). Location-based prioritation of surigao municipalities using probabilistic seismic hazard

analysis (PSHA) and geographical information systems (GIS). 9th IEEE International Conference Humanoid, Nanotechnology, Information Technology Communication and Control, Environment and Management (HNICEM). Manila: The Institute of Electrical and Electronics Engineers Inc. (IEEE) – Philippine Section.

[14] Air Field Manuals. (1999, October 27). Materials Testing.

[15] Wagner, A. (1957). The Use of the Unified Soil Classification System by the Bureau of Réclamation. Denver, Colorado: Bureau of Reclamation. Retrieved February 25, 2018, from https://www.issmge.org/uploads/publications/1/41/1957_01_0030.pdf

[16] Statsoft. (2018, February). k-Nearest Neighbors. Retrieved from Statsoft: http://www.statsoft.com/Textbook/k-Nearest-Neighbors

[17] Ertugrul, O., & Tagluk, M. (2017). A novel version of k nearest neighbor: Dependent nearest neighbor. Applied Soft Computing, 480–490.