Paper ID: CE 0310

Prediction of Bearing Capacity of Pile Using Machine Learning Approach

M. M. H. Galib[1], M. M. Billah[2], S. A. Mofiz[3]

[1]Department of Civil Engineering, RUET, Bangladesh (mehedihassangalib@gmail.com)
[2]Department of Civil Engineering, RUET, Bangladesh (mostain.billah113@gmail.com)
[3]Department of Civil Engineering, RUET, Bangladesh (samofiz@gmail.com)

Abstract

While constructing any geotechnical structures, estimation of the bearing capacity of the piles is of great concern. The traditional techniques used for the calculation of the bearing capacity of piles are time-consuming and costly. Rather some of them are empirical and don't consider all the variables. Therefore, many researchers nowadays are working on various machine-learning approaches to model all the variables for predicting the pile-bearing capacity with greater accuracy. In this study, cohesion, friction angle, specific weight of soil, pile-soil friction angle, flap number¸ pile area, and pile length have been considered as the contributing factors of pile-bearing capacity. Hence various machine learning models have been built for predicting the bearing capacity of the pile. Datasets have been collected from various regions for modeling the heterogeneous nature of the soil and avoiding the overfitting issue to make the model more generalized. From this study, the best R^2 values have been obtained as 0.98 for steel piles from Extreme Gradient Boosting and 0.95 for concrete piles from Random Forest Regression. So, it has been proposed that Extreme Gradient Boosting and Random Forest Regression models can be used for predicting the pile-bearing capacity of steel and concrete pile respectively.

Keywords: *Bearing capacity; Machine learning; Prediction; Extreme Gradient Boosting; Random Forest regression*

1 Introduction

In the field of geotechnical engineering, piles are commonly used as a foundation element to support various types of structures such as buildings, bridges, and offshore platforms. In pile foundation engineering, pile-bearing capacity is an important parameter that must be evaluated (Drusa et al., 2016). Nowadays, several approaches have been developed in order to provide alternative techniques and methods which involve numerical, experimental, and analytical approaches for estimating the bearing capacity of piles (Shahin, 2010). The SPT is commonly used to evaluate pile-bearing capacity (Bouafia et al., 2002). Several propositions based on SPT results have been proposed to predict pile-bearing capacity, including empirical equations derived from different research (Bazaraa and Kurkur, 1986) and (Shariatmadari et al., 2008). It has been also suggested for adopting an experimental formula to account for the effects of soil type or utilizing the SPT value for sandy soil and the untrained shear strength of soil (Cu) for clayey soil (Architectural Institute of Japan, 2004). However, estimating the bearing capacity of piles using the aforementioned approaches has been proven to be time-consuming and costly (Abu-Farsakh et al., 2004). None of the methodologies that were suggested in the literature produced accurate results (Kordjazi et al., 2014). As a result, the application of machine-learning approaches to predict pile-bearing capacity has grown significantly since the early 1990s (Ikeagwuani, 2021). Their benefits include the ability to deal with large amounts of data and to explore complex and highly nonlinear relationships between various parameters. A recent study has proposed using a random forest algorithm to predict the bearing capacity of single piles in different soil types (Wu et al., 2020). Similarly, a support vector machine algorithm was more accurate than conventional empirical methods in predicting the pile-bearing capacity (Singh et al., 2019). Again a K-nearest neighbor algorithm has been used to predict the bearing capacity of piles in clayey soils which provided accurate predictions of pile-bearing capacity with an average relative error of 9.1% (Jiang et al., 2019).

Linear regression could provide accurate predictions with an average relative error of 6.5% (Liu et al., 2018). In addition, a machine learning model based on gradient boosting regression to predict the bearing capacity of pile foundations using data from a field test had better performance in predicting the pile capacity compared to other machine learning models (Kardani et al., 2020).

2 Methodology

2.1 Data Preprocessing

The dataset which has been used in this study contains 8 variables and 100 observations of steel and concrete piles. The first 7 variables have been used as explanatory variables and pile capacity has been used as response variables. The statistical properties of the dataset have been shown in Table 1 and in Table 2 for concrete and steel piles respectively.

Table 1. Statistical properties for concrete pile

Statistical Properties	Variables of Dataset							
	Average Cohesion (kN/m^2)	Average Friction angle (°)	Average soil Specific weight (kN/m^3)	Average Pile-Soil friction angle (°)	Flap Number	Pile Area (m^2)	Pile Length (m)	Pile Capacity (kN)
Mean	39.75	22.15	9.65	14.09	229.74	0.17	22.96	2388.36
Standard Deviation	47.09	10.50	1.72	1.25	370.75	0.11	4.87	677.19
Correlation with Pile Capacity	-0.22	0.34	0.1	0.18	0.34	0.19	0.23	1.00

Table 2. Statistical properties for steel pile

Statistical Properties	Variables of Dataset							
	Average Cohesion (kN/m^2)	Average Friction angle (°)	Average soil Specific weight (kN/m^3)	Average Pile-Soil friction angle(°)	Flap Number	Pile Area (m^2)	Pile Length (m)	Pile Capacity (kN)
Mean	29.94	28.32	10.71	13.47	542.53	0.54	24.08	3171.25
Standard Deviation	17.34	8.56	2.05	2.12	542.89	0.50	12.71	1538.43
Correlation with Pile Capacity	0.13	0.35	0.21	-0.07	0.45	0.50	0.07	1.00

In the data preprocessing phase, the dataset has been checked whether it has been fit to be used in machine learning algorithms or if some corrections have been required. Missing values and irrelevant observations have also been checked. Although there were no missing values, some outliers were present in the dataset. These outliers have been removed so that the algorithms would not consider any unusual cases in the models.

Since the dataset contains two types of piles, the whole dataset has been segregated according to the materials of the pile: steel pile and concrete pile. Henceforth each dataset has been split into two parts: the training dataset and the test dataset. 80 percent of the observations have been used as train set for training the models while the remaining 20 percent have been used as a test set for evaluating the prediction accuracy of the models.

2.2 Machine Learning Algorithms

2.2.1 Random Forest Regression

Random forest regressor is a supervised machine learning algorithm where decision trees have been used as base learners. It involves an ensemble learning method for prediction. The number of decision trees in a random forest regressor generally depends on the designers' choice through a trial & error process. The main feature of this algorithm is the resampling of the training dataset along with replacement. The resampled datasets have been used to train each decision tree model incorporated in random forest. Then the average of all decision tree models has been considered as the output of the random forest model. This procedure has been demonstrated through a flow chart in figure 1. Hence resampling along with replacement helps to improve the accuracy of prediction by building a more generalized model and overcoming the overfitting issue.

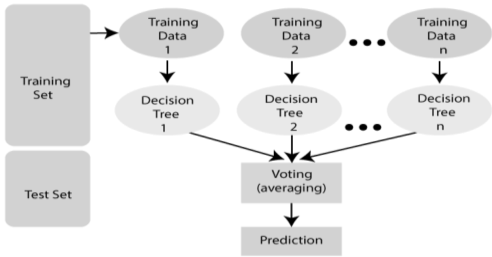

Figure 1. Typical structure of Random Forest algorithm.

While building a random forest algorithm, hyperparameter tuning has also been carried out to observe the variation in the prediction accuracy of the model. Hyperparameters such as the number of decision tree models and the maximum number of leaf nodes each decision tree can contain have been considered through a trial-and-error process.

2.2.2 Extreme Gradient Boosting Regression

Extreme Gradient Boosting (XGBoost) Regression is a supervised machine learning algorithm where gradient boosting and decision trees have been used as base learners. It shares similarities with Random Forest, yet offers unique advantages. It utilizes ensemble learning techniques for prediction. However, instead of using a fixed number of decision trees, the number of trees has been dynamically adjusted during the training process based on performance, preventing overfitting and optimizing the model's generalization ability. A notable feature is its gradient boosting framework, which involves iteratively building decision trees to correct errors made by previous models shown in figure 2. Applying this iterative process, prediction accuracy has been gradually improved by focusing on the instances that are difficult to predict. To further enhance performance, in XGBoost advanced regularization techniques, such as L1 and L2 regularization have been used to control model complexity and mitigate overfitting. During the training phase, hyperparameter tuning has been done in optimizing the prediction accuracy. Parameters such as the number of decision trees, maximum tree depth, and learning rate have been carefully adjusted through a trial-and-error process to find the optimal configuration.

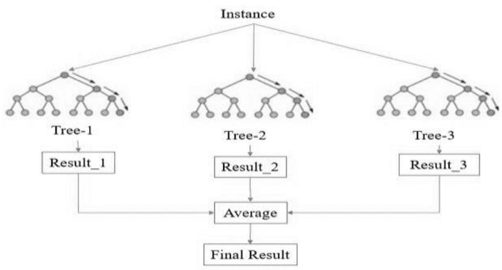

Figure 2. Typical structure of XGBoost Regression.

XGBoost Regression provides not only accurate regression predictions but also valuable insights into feature importance. By analyzing the contribution of each feature to the model's predictions, the key variables that influence the target variable have been identified.

2.3 Evaluation of the Model

Once the model has been trained, it has been required to evaluate the model using the test dataset. There are lots of techniques for carrying out performance measurement as well as error metrics to evaluate the models. In this study, coefficient of determination (R^2), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE) have been used as error metrics.

Formula of (R^2):

$$R^2 = 1 - \frac{ss_{res}}{ss_{tot}} \tag{1}$$

Here,
SS_{res} = Residuals sum of squares
SS_{tot} = Total sum of squares

Formula of MAE:

$$\mathrm{M} = \frac{1}{n}\sum_{t=1}^{n}\left|A_t - P_t\right| \tag{2}$$

Formula of MAPE:

$$\mathrm{M} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - P_t}{A_t}\right| \tag{3}$$

Formula of RMSE:

$$\mathrm{M} = \sqrt{\sum_{t=1}^{n}\frac{(A_t - P_t)^2}{n}} \tag{4}$$

Here,
M = Error metrics
A_t = Actual value
P_t = Predicted value
n = Number of observations

3 Result and Discussion

3.1 Random Forest Regression

After training the models with the training dataset, the models have been evaluated using the test dataset for steel and concrete pile separately. The error metrics for both steel pile and concrete pile have been shown in Table 3 and Table 4 respectively. From the analysis of the steel pile, the best R^2 value has been found 0.97 for the hyperparameter set of (500, 100). Here 500 and 100 are the number of decision tree models and the maximum number of leaf nodes each decision tree can contain respectively. The other error metrics of the same model such as MAE, MAPE, and RMSE have been found 261.06, 9.62%, and 324.62 respectively, which are also comparatively better than the other models of different hyperparameter sets.

Table 3. Error metrics for Random Forest algorithm (Steel Pile)

Hyper-parameters	Error Metrics			
	R^2	MAE	MAPE(%)	RMSE
500, 100	.97	261.06	9.62	324.62
1000, 10	.96	334.14	12.08	389.48
100, 100	.96	282.78	10.10	357.11

Table 4. Error metrics for Random Forest algorithm (Concrete Pile)

Hyper-parameters	Error Metrics			
	R^2	MAE	MAPE(%)	RMSE
500, 100	.95	546.08	16.85	627.21
1000, 10	.95	530.68	16.12	626.40
100, 100	.92	605.38	19.24	688.34

For concrete piles, the best R^2 value has been found 0.95 for hyperparameter sets (500, 100) and (1000, 10). But if the other error metrics have been taken into account, it has been seen that the model with hyperparameters (1000, 10) has been comparatively better as the MAPE for this model has been found 16.12%.

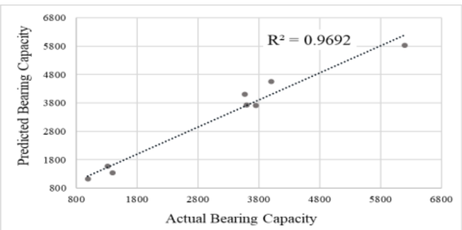

Figure 3. R^2 for steel pile (500, 100)

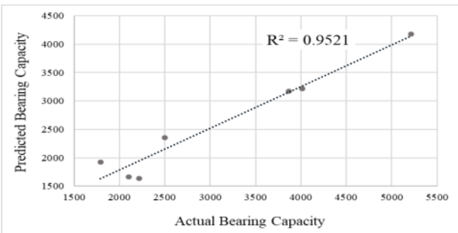

Figure 4. R^2 for concrete pile (500, 100)

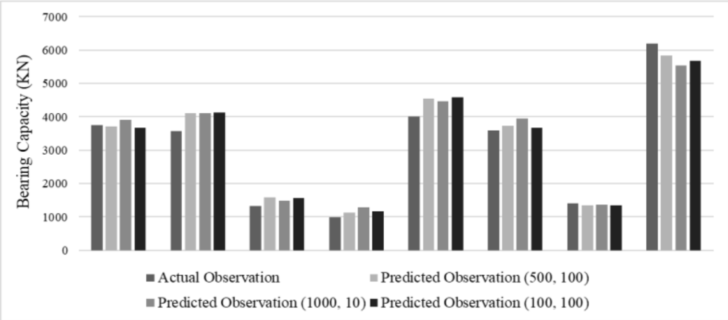

Figure 5. Actual vs Predicted Bearing Capacity for Random Forest (Steel Pile)

3.2 Extreme Gradient Boosting Regression

According to figure 6 and figure 7, the R^2 values in XGBoost have been found 0.98 and 0.91 for steel and concrete pile respectively, which have been found comparatively better than the two previous models. Though Random Forest and XGBoost both are ensemble learning algorithms based on Decision Tree, XGBoost has performed better than Random Forest due to owing some unique features. Random forest is completely built on random sub-datasets, while the XGBoost sub-datasets completely depend on the previous loss functions. The other error metrics such as MAE, MAPE, and RMSE are 157.59, 5%, and 212.36 respectively for steel piles and 325.73, 15.38%, and 358.24 respectively for concrete piles. All the results obtained from Extreme Gradient Boosting Regression have been shown in Table 5.

Table 5. Error metrics for Extreme Gradient Boosting Regression

Error Metrics	**Steel Pile**	**Concrete Pile**
R^2	0.98	0.91
MAE	157.59	325.73
MAPE (%)	5.00	15.38
RMSE	212.36	358.24

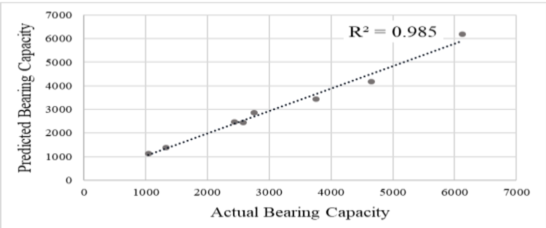

Figure 6. R^2 for Steel Pile.

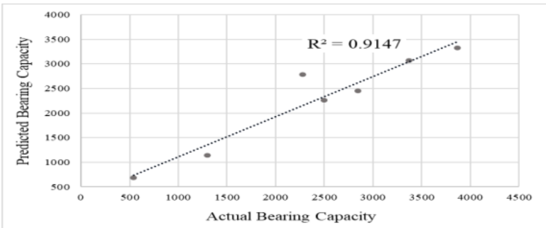

Figure 7. R^2 for Concrete Pile

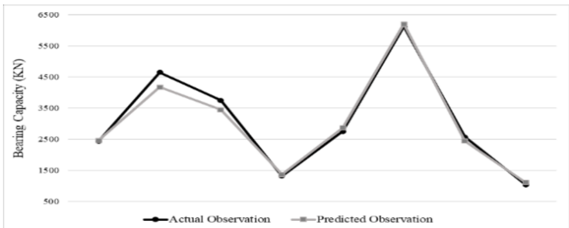

Figure 8. Actual vs Predicted Bearing Capacity for Steel pile.

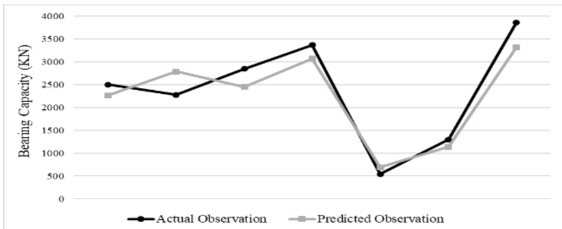

Figure 9. Actual vs Predicted Bearing Capacity for Concrete Pile.

2.3 Validation of the models

To assess the accuracy and reliability of these models, a precise validation approach is conducted.

Table 6. Comparison between R^2 values

Model	Findings of our study (R^2 value)	Findings of previous study (R^2 value)	Reference
Random Forest Regression	0.970	0.919	Kardani et al. (2020)
Extreme Gradient Boosting Regression	0.980	0.975	Kardani et al. (2020)

By comparing the values with the values of R^2 conducted by Kardani et al. (2020) in Table 6, it is found that the R^2 values obtained from the models are consistent with the compared values.

4 Conclusion

From the study, it has been found that Extreme Gradient Boosting has performed quite well in predicting the bearing capacity of piles as compared to Random Forest in the case of steel piles. And it has also been found that machine learning algorithms can model the characteristics of steel piles better as compared to concrete piles. Hence the R^2 value has been found 0.98 for steel piles, while the R^2 value of concrete piles has been found only 0.91 in Extreme Gradient Boosting. The performance of Extreme Gradient Boosting in predicting the bearing capacity of steel piles has also been found satisfactory compared to the concrete pile. Therefore, the MAPE for steel piles has been found only 5%, whereas the MAPE for concrete piles has been found 15.38%. For concrete piles, Random Forest with hyperparameter sets (500, 100) and (1000, 10) have performed comparatively better than the Extreme Gradient Boosting. The R^2 value from these models has been found 0.95, which is satisfactory as compared to the performance of Extreme Gradient Boosting in the case of concrete piles. Since the accuracy of prediction obtained from Extreme Gradient Boosting and Random Forest is satisfactory for steel and concrete pile respectively, these algorithms can be used in the practical field to predict the bearing capacity of piles. And for improving the prediction accuracy, more advanced hyperparameter tuning can be carried out. Furthermore, some other advanced machine learning models along with deep learning models can be used to improve the accuracy in predicting the bearing capacity of pile.

References

Abu-Farsakh, M.Y.; Titi, H.H. (2004). Assessment of Direct Cone Penetration Test Methods for Predicting the Ultimate Capacity of Friction Driven Piles. J. Geotech. Geoenvironmental Eng.130, 935–944.

Architectural Institute of Japan. (2004). Recommendations for Design of Building Foundation; Architectural Institute of Japan: Tokyo, Japan.

Bazaraa, A.R.; Kurkur, M.M. (1986). N-values used to predict settlements of piles in Egypt. In Proceedings of In Situ'86; New York, American Society of Civil Engineers: New York, NY, USA; pp. 462–474.

Bouafia, A.; Derbala, A. (2002). Assessment of SPT-based method of pile bearing capacity–analysis of a database. In Proceedings of the International Workshop on Foundation Design Codes and Soil Investigation in View of International Harmonization and Performance-Based Design, Tokyo, Japan, pp. 369–374.

Ikeagwuani, C.C. (2021). Estimation of modified expansive soil CBR with multivariate adaptive regression vsplines, random forest and gradient boosting machine. Innov. Infrastruct. Solut. 6, 199.

Jiang, J., Li, W., & Wang, J. (2019). Prediction of pile bearing capacity based on K-nearest neighbor algorithm Measurement, 136, 163-173.

Kardani, N.; Zhou, A.; Nazem, M.; Shen, S.-L. (2020) Estimation of Bearing Capacity of Piles in Cohesionless Soil Using Optimised Machine Learning Approaches. Geotech. Geol. Eng. 2020, 38, 2271–2291.

Kordjazi, A.; Pooya Nejad, F.; Jaksa, M.B. (2014). Prediction of ultimate axial load-carrying capacity of piles using a support vector machine based on CPT data. Comput. Geotech. 55, 91–102

Liu, H., Guo, Y., Zhang, X., & Xu, H. (2018). Predicting pile-bearing capacity using linear regression in different soil conditions. Advances in Civil Engineering, 1-8.

M. Drusa, F. Gago, and J. Vlˇcek. (2016). Contribution to estimating bearing capacity of pile in clayey soils, Civil and Environmental Engineering, vol. 12, no. 2, pp. 128–136.

Shahin, M.A. (2010). Intelligent computing for modeling axial capacity of pile foundations. Can. Geotech. J. 47, 230–243.

Shariatmadari, N.; Eslami, A.A.; Karim, P.F.M. (2008). Bearing capacity of driven piles in sands from SPT applied to 60 case histories. Iran. J. Sci. Technol. Trans. B Eng. 32, 125–140.

Singh, S., Puri, V., & Kumar, P. (2019). Support vector machine-based prediction of pile bearing capacity. Geotechnical and Geological Engineering, 37(6), 4989-4998.

Wu, X., Hu, H., Fan, X., & Wu, W. (2020). A random forest model for predicting bearing capacity of single piles in different soil types. Engineering Applications of Artificial Intelligence, 94, 103806.