A report on

Supervised Learning Method for Modelling Unconfined Compressive Strength Prediction of Lime and Calcium Stabilised Soil

Submitted By

Nandita Suresh Kamath 2018B4A20868P

Under the Guidance of

Dr. Sayantan Chakraborty

Birla Institute of Technology and Science, Pilani

1. Introduction to Soil Stabilisation

Soil stabilisation is the method of increasing the strength and durability and reducing the compressibility and permeability of natural soil by blending it with additives such as admixtures or stabilisers. Any natural soil, cementing mixture, chemical, binder materials or by products of cement manufacture can be used to undesirable soil to improve one or more of its properties. It helps civil engineers alter essential soil properties to maximise its the performance. This method can not only retain moisture content, improve soil particle cohesion, catalyse the cementing of soil particles and prevent the soils affinity to water but also combat the high swelling nature of clay-soil. [1] It can also help increase overall soil strength, durability stiffness and reduce plasticity of soil. This paper deals with the complete analysis of the use of lime and cement to improve of soil properties.

2. Lime Stabilisation

2.1 Introduction

Lime stabilisation has been used for high plasticity soils such as fine clay or silty clay that have a high expanding nature. [2] [3]Clay rich soils when treated with lime results in a decrease of liquid limit, maximum dry density and plasticity index, and an increase in the optimum moisture content, strength and shrinkage limit [4]. It is used to construct base layers pavements and works as water resistant barrier that prevents penetration of rain water in warm, humid regions.

2.1 Principles of Lime Stabilisation

1. Drying of soil - The stabilisation process starts with the drying action by quicklime when poured on clayey soil. It chemically reacts with the free water in soil to create calcium hydroxide. This process is highly exothermic and the heat generated further evaporates additional water present in soil and reduces the soil's moisture holding capacity.

2. Lime modification – Hydrated lime modifies the chemical makeup of the clay that makes the soil agglomerate and flocculate [5]. This happens as divalent cations like calcium ions of lime migrate to the surface of the clay particles and displace water and other monovalent cations. This exchange of cations makes the soil becomes granular and workable. The plasticity index of the soil decreases dramatically, along with swelling and shrinking.

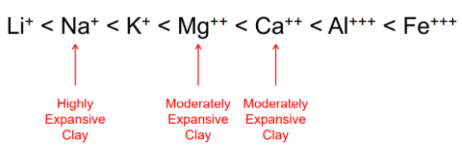

Fig 1 – Order of preference for cation exchange

3. Lime Stabilisation - Quicklime also makes soil water alkaline and helps clay release Si and Al is which combines with Calcium to give CSH and CAH. When lime and water are added, the soil becomes alkaline and this helps the clay particles to break down. Silica and alumina ions are released and react with calcium ions to form calcium-silicate-hydrates (CSH) and calcium-aluminate-hydrates (CAH) [6] [7]. CSA and CAH form the matrix that contributes to the strength of lime-stabilised soil layers. As this matrix forms, the soil is transformed to a sandy, granular material to a hard, relatively impermeable layer with significant load bearing

capacity. The matrix formed is permanent durable, and significantly impermeable, producing a structural layer that is both strong and flexible. [8] CSA and CAH form a mesh that causes the increased strength of lime stabilised soil. This way, the clayey soil is transferred to a granular soil with high load bearing capacity. This pozzolanic reaction increases the strength gradually.

$$Ca^{2+} + OH^{-} + \text{Soluble Clay Silica} \rightarrow \text{Calcium Silicate Hydrate (CSH)}$$

$$Ca^{2+} + OH^{-} + \text{Soluble Clay Alumina} \rightarrow \text{Calcium Alumina Hydrate (CAH)}$$

[9] [1]

Lastly it cements clay particles together by creating $CaCO_3$ from CaO and atmospheric CO_2

$$Ca(OH)_2 + CO_2 \rightarrow CaCO_3 + H_2O$$

[8] [10]

2.3 Types of Lime Used

1. Quick lime (CaO)
2. Hydrated lime [$Ca(OH)_2$]
3. Dolomite lime (CaO+MgO or $Mg(OH)_2$) [11] [12]

Quicklime which is high in calcium is the most commonly used lime and is better than slaked or hydrated lime as it contains a higher amount of free lime, is denser due to high calcium content, less contaminated with dust and produces more heat which evaporates the water and improves durability. Hydrated dolomite lime which has a high concentration of magnesium has less water affinity and produces low amounts of heat during mixing and hence, is rarely used. [13]

2.5 Process of Lime Stabilisation

The process of lime stabilisation starts with preparing the soil and breaking it down to a uniform, lump-free mixture. The particle size of this pulverized mixture should not be more than 5 mm to ensure effective mixing. Lime is added as powder or slurry to clay-soil and mixed manually or with the help of machines. Slurry lime ensure effective mixing. When dry powder is added to soil, the soil needs to be made moist in advance. The mixture is allowed to rest for lime-soil reaction to take place. The rest of the lime is added and mixed again thoroughly. At this stage, additives such as cement and fly ash or sulphates such as Sodium metasilicate, Sodium hydroxide and Sodium Sulphate and carbonates maybe added to improve strength. The obtain maximum strength from the mix, it is spread to required grade and compacted at optimum moisture content to achieve highest dry density. The compressed mixture is cured in water. The strength of mixture increases rapidly during the initial few days up to several years. Environmental temperature and humidity of the surroundings also effects the strength. Field tests are conducted to check on dry density, water content and strength of stabilised soil. [10]

3. <u>Cement Stabilisation</u>

3.1 Introduction

Cement stabilisation has been used for granular soils with high organic content and deleterious salts with high sulphate and carbonate content. It is used to construct base layers pavements and works as water resistant barrier that prevents penetration of rain water in warm, humid regions.

3.2 Principles of Cement stabilisation

The principles of cement stabilised soil are very same to that of lime stabilised soil. Cation exchange of calcium ions with other ions is followed by particle restructuring during flocculation of soil particles. This causes a pozzolanic reaction between cement and soil followed by carbonate cementation. An important theory special to soil-cement stabilisation is the formation of a gel/crystal like film surrounding cement and soil particles at the time of hydration. With the progression of the hydration, these crystals grow to form a network that contributes to improved strength. Cement stabilisation is independent of the minerals in soil but heavily dependent upon the amount of water used during hydration. The hydration also makes the soil-cement mixture more workable. [14]

3.3 Process of cement stabilisation

Pulverised soil is thoroughly mixed with Portland cement and water. This soil-cement mixture is compressed to get a strong and durable material during cement hydration. The compaction process is continued by adding more cement in the voids of the soil particles and hence reduce the void ration in the mixture. Hydration allows for this mixture to gain further strength and also increase its unit weight, shear strength and bearing capacity and reduced permeability of soil.

3.4 Types of Cement Used

Apart from ordinary Portland cement, blast furnace cement, sulphate resistant cement and high alumina cement may also be used depending on the type of soil to be treated and desired final strength.

4. Advances in Machine Learning in Geotechnical Engineering

A. H. Alavi used Artificial Neural Networks to predict the values of maximum dry unit weight and optimum moisture content of soil-stabilizer mix based on liquid limit, plastic limit, linear shrinkage, percentage sand, percentage fines, lime content, cement content, and asphalt content. [15] Das SK predicted the unconfined compressive strength and MDUW of cement stabilized soil where different classes of ANN and SVM were implemented [16]. Furthermore, Mozumder and Laskar reported compared the results of ANN and MLR in predicting UCS value of geopolymer stabilized clayey soil based on several predictors, including LL, plasticity index (PI), percentage stabilizers, molar concentration of alkali activator, ratio of alkali to binder, ratio of Na/Al, and ratio of Si/Al. [17] Tinoco researched the applicability of various machine learning models in prediction of UCS value for jet grouting material was presented. [18]

5. Supervised Machine Learning Predictive Model

5.1 Problem Statement

Create a ML model which can predict the strength of lime and cement stabilised soil

- Target Variable: Unconfined Compressive Strength
    - Predictors: Organic content, sand, clay, silt, gravel

As the target variable is given in the dataset, a supervised ML Regression model is implemented.

5.2 Current Case Study and Database Used

Machine Learning is being used to develop predictive models in various fields to predict different aspects of a problem. This study utilizes the basic concepts of a supervised ML model and uses soil properties such as silt, gravel, sand, clay and organic content as predictor variables to forecast unconfined compressive strength and construct a predictive model to help civil engineers save time and money on cumbersome experiments in the soil stabilisation process. The sample dataset consisting of the type and amount of its constituents and its strength at various stages of curing were compiled from the published literature. This database was then split into a training and testing set to apply different ML models and the ones providing highest accuracy were optimized to solve this supervised ML Regression problem.

5.3 Model development

A predictive ML model is developed after compilation, exploration and cleaning of the data. The language used to develop this model is Python and the online notebook used to code and deploy the model is Kaggle. The project is approached by defining the problem statement, analysing the dataset in python using data exploration tools, identifying predictor and target variables, dividing the sample space into training and testing set and trying multiple regression algorithms to deploy the best model in production. The dataset is applied on numerous models such as Multiple Linear Regression model, Decision Tree, AdaBoost, Support Vector Machine and Random Forest to find a suitable model with maximum accuracy. [19] As seen in Figure 1, the lime dataset is divided into training and testing sets and ML models are trained on its input parameters to build a predictive model.

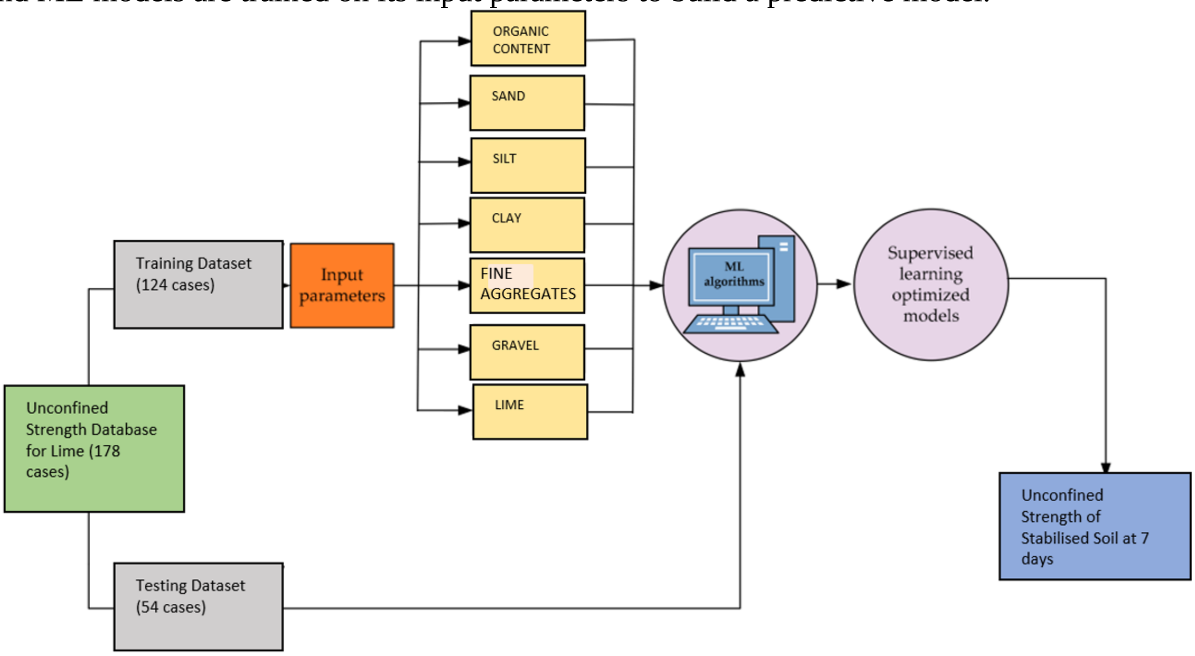

Fig 2 – Framework of the proposed ML model (Lime)

5.4 Work flow

1. Data Catalogue Compilation:

Table 1 – Sample of Unconfined Compressive Strength Catalogue (Ex - Lime)

1	Name of Soil	Classification (USCS	Organic.Content	Sand	Silt	Clay	Clay2	Gravel	LL-LIQUID LIMIT	PL-PLASTICITY LIMIT	PI-PLASTICITY INDEX	Lime	UCS.psi	UCS_std_Soaked_7Days	UCS_std_Unsoaked_7Days	UCS_std_Soaked_28Days	UCS_std_Unsoaked_28Days	Source
2	Alluvial	CH	1.33	2	29	69	98	0	72	26	46	0	0	0	40	0	40	Davidson-Iowa-State
3		CH	1.33	2	29	69	98	0	72	26	46	2	25.3	25.3	56.55	28.27	58.04	Davidson-Iowa-State
4		CH	1.33	2	29	69	98	0	72	26	46	4	107.16	107.16	122.04	123.53	157.76	Davidson-Iowa-State
5		CH	1.33	2	29	69	98	0	72	26	46	8	169.67	169.67	171.162	303.627	308.09	Davidson-Iowa-State
6		CH	1.33	2	29	69	98	0	72	26	46	12	174.14	174.139	181.58	320	320	Davidson-Iowa-State

The data entries are compiled into two separate excel files, one for Lime and another for cement. The lime dataset with 178 entries and the cement dataset with 58 entries are compiled on an excel sheet and read into python as a CSV file. [20]

2. Basic Data Exploration:
Data analysis and exploration are performed to gauge the spread of the data. Each data point is analysed for its type, and whether it affects the values of the target variable. Python library functions are used to determine the size of the dataset and the number of rows and columns. The exact statistical details of the data such as minimum, maximum, average and standard deviation is found. Each variable is identified as categorical or continuous and the missing values are flagged.

Table 2 - Statistics of important variables for the dataset used in the development of the supervised learning model

	Organic.Content	Sand	Silt	Clay	Clay2	Gravel	LL-LIQUID LIMIT	PL-PLASTICITY LIMIT	PI-PLASTICITY INDEX	Lime	UCS.psi
count	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000
mean	0.993202	16.410112	41.022472	37.646067	78.550562	5.078652	45.258427	21.393258	22.168539	5.977528	104.787079
std	1.467133	15.940756	19.340527	17.304959	23.388640	13.483805	13.840927	7.130927	13.013887	4.108631	80.344481
min	0.000000	0.000000	5.000000	0.000000	15.000000	0.000000	20.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	2.000000	30.000000	29.000000	63.000000	0.000000	37.000000	18.000000	15.000000	2.000000	34.550000
50%	0.170000	12.000000	37.000000	39.000000	84.000000	0.000000	43.000000	21.000000	21.000000	6.000000	106.775000
75%	1.600000	29.750000	57.000000	46.000000	98.000000	2.000000	52.000000	26.000000	28.000000	10.000000	158.900000
max	4.770000	65.000000	81.000000	75.000000	100.000000	70.000000	76.000000	34.000000	54.000000	14.000000	338.460000

3. Data Mining:
An important step in predictive modelling is cleaning the sample set. Only the input variables that represent the system and effect the target variable need to be included, and the rest need to be removed. This is done by filtering the CSV file and removing unnecessary columns and rows that will not contribute to the ML model such as serial number, soil type and source.

   The variables retained are Organic Content, Sand, Silt, Clay, Clay 2(Summation of silt and clay), Gravel, Lime and the target variable UCS (Unconfined compressive strength)

4. Visual Exploratory Data Analysis:
Continuous Variables are plotted on histograms. The ideal outcome is a bell curve however, in reality, datasets are rarely perfect. Hence, the variables are accepted as long as there is good distribution across its range.

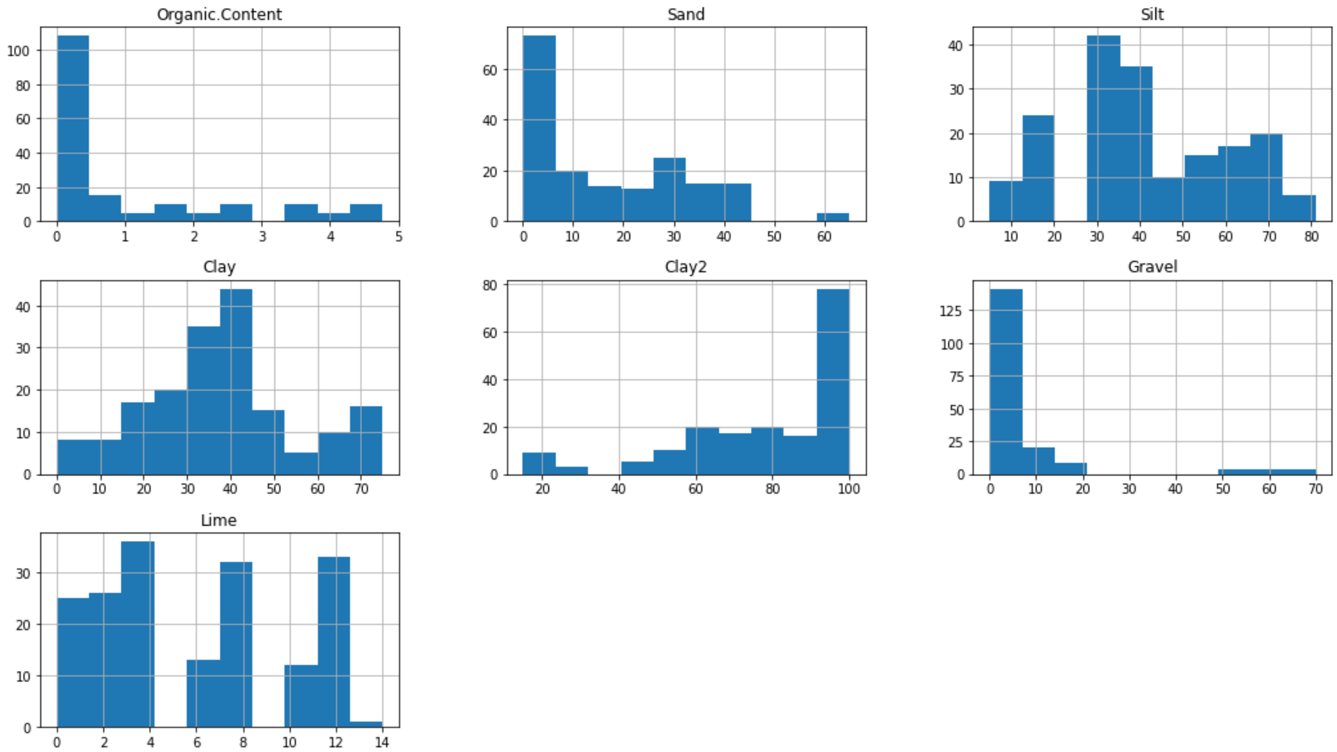

Fig 3 - Distribution of the predictor variables

More importantly, the target variable is analysed and plotted on a histogram. The data distribution of the target variable is satisfactory to proceed further. There are sufficient number of rows for each type of values to learn from.

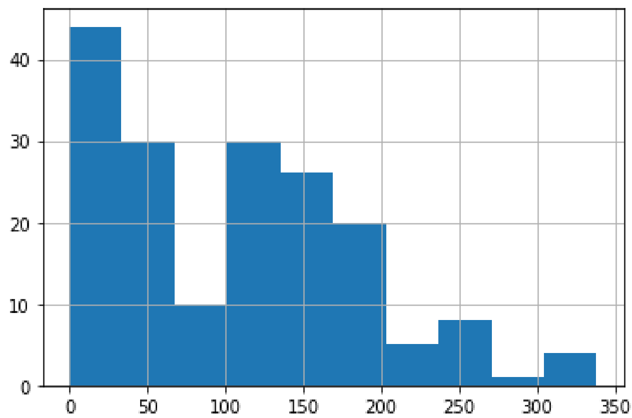

Fig 4 – Distribution of the target variable

5. Relationship exploration:
   Since the target and predictor variable is continuous in nature, a scatter plot is used to visualise their relationship and measure the strength of relation using Pearson's correlation value.

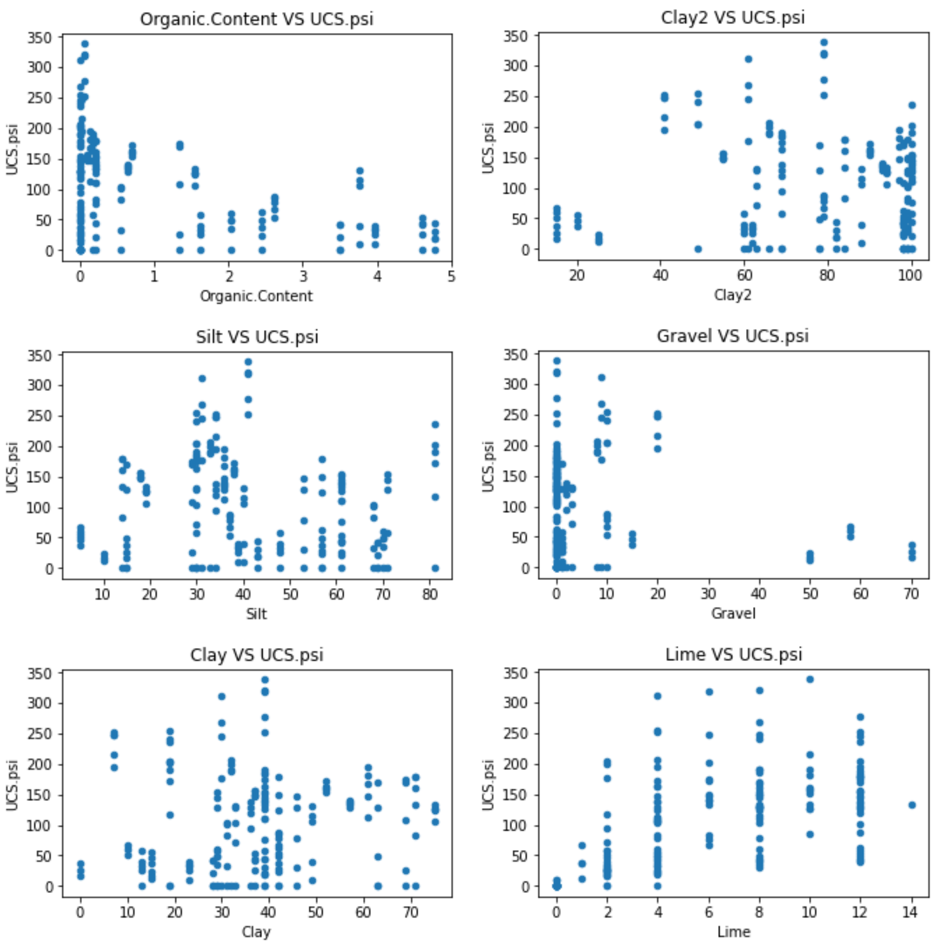

Fig 5 - Scatter Plot depicting the trend between each input variable and UCS at 7 days

Variables that are positively correlated and are directly proportional to each other have an increasing trend and variables that are negatively correlated and are inversely proportional to each other have a decreasing trend.

6. Statistical Feature Selection using Pearson's Correlation value:
   The correlation value can be calculated for any two numeric variables. A negative correlation indicates an inversely proportional relation (downward trend) and a correlation between 0 and 1 indicates a directly proportional relation (upward trend). By studying the correlations between Target variable and all other predictor variables, the intensity of the effects of the predictors can be checked.

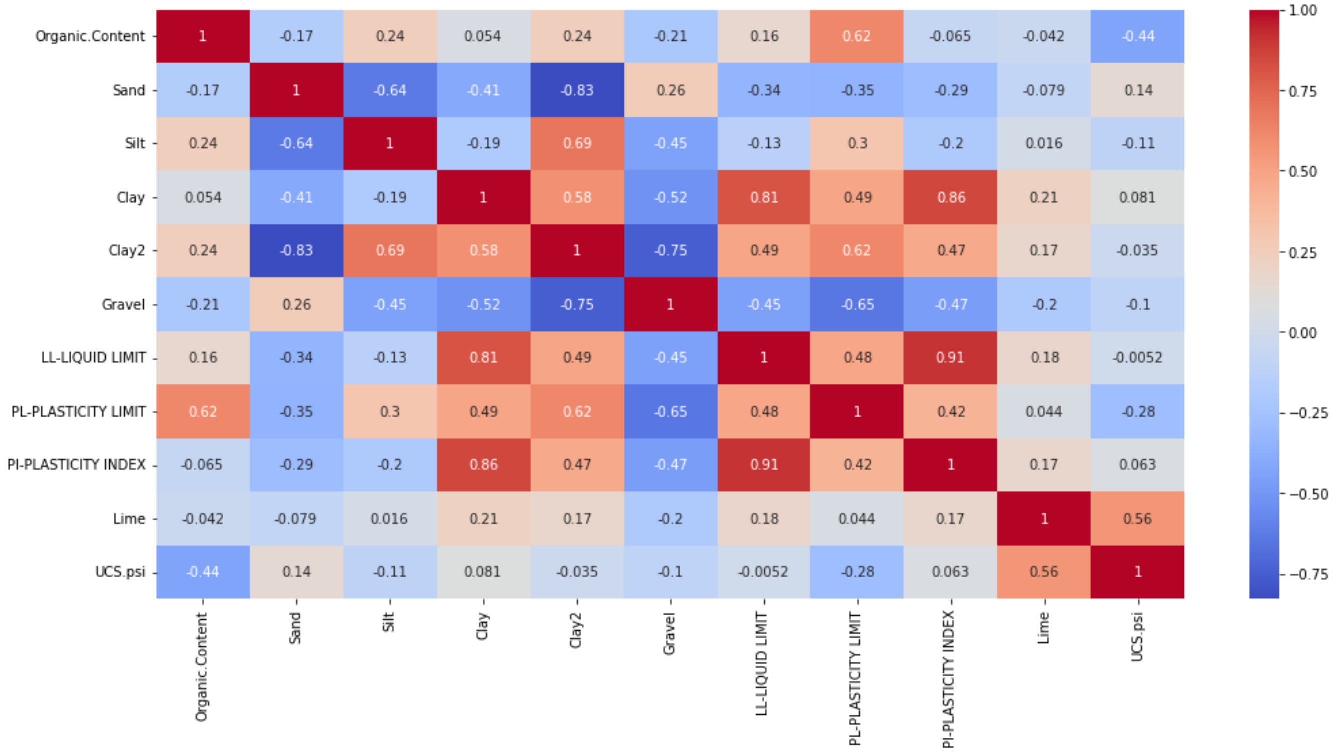

Fig 6 – Heat map and correlation matrix of the variables

7. Dividing the dataset into Training and Testing sample:
   The existing dataset is shuffled, selected at random and divided 2 sets; Training and Testing with a 70-30% ratio. The training set is used to build the model and the testing set is used to validate it.

8. Model Selection and Optimization:
   Linear Regression, Decision Tree, L2 (Ridge) Regression, Support Vector Machine (Linear and RBF Kernel), ANN, Random Forest, Gradient Boosting and AdaBoost are the models used. The model providing with the highest accuracy is selected and optimized. Later, the model is saved and the predicted values are analysed against the original values of the training set.

6 Results and Conclusion

The predictive model given above is applied to both the available datasets, lime and cement. After analysing the results and accuracy provided by each model, the one with highest accuracy is selected and optimized. Gradient Boosting model gave the highest accuracy of **83.73%** for the lime dataset and for the cement dataset, the Ridge regression model gave the highest accuracy of **86.25%.** The accuracy of these ML models is satisfactory with respect to our dataset and the comparison of the values predicted by the model with the original laboratory findings supports the applicability of this model for stabilised soil strength prediction. New input variables can be added to the algorithm as mentioned in the appendix to find new strength values.

7 Appendix

1. Importing all relevant libraries

```
import numpy as np
import pandas as pd
import seaborn as sb
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression, Ridge
from sklearn.svm import LinearSVR, SVR
from sklearn.tree import DecisionTreeRegressor
from sklearn.neural_network import MLPRegressor
from sklearn.ensemble import RandomForestRegressor,
GradientBoostingRegressor, AdaBoostRegressor
from sklearn.model_selection import GridSearchCV
```

2. Reading the CSV file into python

```
datas = pd.read_csv('../input/limedataset/LimeDataSet.csv')
datas
```

3. Extracting number of rows and column

```
datas.shape
```

4. Extracting number of data points

```
datas.size
```

5. Extracting first 5 rows of the dataset

```
datas.head()
```

6. Extracting summarized information regarding the number of variables and type of variable(float, int or str)

```
datas.info()
```

7. Extracting statistical information such as average and standard deviation for each variable

```
datas.describe()
```

8. Extracting unique value for each variable to determine if the variable is continuous or categorical

```
datas.nunique()
```

9. Extracting number of missing values for each variable

```
datas.isnull().sum()
```

10. Data Cleaning – Filtering the dataset by removing unnecessary columns that don’t affect the target variable

```
datas = datas.iloc[: , :-5]
datas = datas.iloc[: , 2:]
datas
```

11. Visual data exploration – plotting histograms for all continuous predictor and target variables

```
datas.hist(['Organic.Content', 'Sand', 'Silt', 'Clay', 'Clay2','Gravel',
'Lime'], figsize=(18,10))
datas['UCS.psi'].hist()
```

12. Statistical feature selection using correlation values to plot heat map

```
plt.figure(figsize=[17,8])
sb.heatmap(datas.corr(),annot=True, cmap='coolwarm')
```

13. Plotting correlation matrix

```
ContinuousCols=['Organic.Content', 'Sand', 'Silt','Clay',
'Clay2','Gravel','Lime','UCS.psi']
CorrelationData=datas[ContinuousCols].corr()
CorrelationData
```

14. Scatter plot for each predictor vs the target variable

```
ContinuousCols=['Organic.Content','Sand','Silt','Clay','Clay2','Gravel','Li
me']
for predictor in ContinuousCols:
datas.plot.scatter(x=predictor, y='UCS.psi', figsize=(5,3),
title=predictor+" VS "+ 'UCS.psi')
```

15. Selecting all suitable predictor variables for ML model

```
SelectedColumns=['Organic.Content','Sand','Silt','Clay','Clay2','Gravel','L
ime']
DataForML=datas[SelectedColumns]
DataForML
```

16. Treating all the nominal variables at once using dummy variables and adding Target Variable to the data

```
DataForML_Numeric=pd.get_dummies(DataForML)
DataForML_Numeric['UCS.psi']=datas['UCS.psi']
DataForML_Numeric.head()
```

17. Separate Target Variable and Predictor Variables and splitting data into training and testing set

```
TargetVariable='UCS.psi'
Predictors=['Organic.Content','Sand','Silt','Clay','Clay2','Gravel','Lime']
X=DataForML_Numeric[Predictors].values
y=DataForML_Numeric[TargetVariable].values
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=428)
```

18. Sanity check for the sampled data before testing models

```
print(X_train.shape)
print(y_train.shape)
print(X_test.shape)
print(y_test.shape)
```

19. Testing various ML models

```
models = {
```

```
    "                      Linear Regression": LinearRegression(),
    "                  L2 (Ridge) Regression": Ridge(),
    "Support Vector Machine (Linear Kernel)": LinearSVR(),
    "   Support Vector Machine (RBF Kernel)": SVR(),
    "                          Decision Tree": DecisionTreeRegressor(),
    "                         Neural Network": MLPRegressor(),
    "                          Random Forest": RandomForestRegressor(),
    "                      Gradient Boosting": GradientBoostingRegressor(),
    "                               AdaBoost": AdaBoostRegressor()
}

for name, model in models.items():
    model.fit(X_train, y_train)
    print(name + " trained.")

for name, model in models.items():
    print(name + " R^2: {:.5f}".format(model.score(X_test, y_test)))
```

20. Optimization

```
best_model = GradientBoostingRegressor()
best_model.fit(X_train, y_train)

print("Model R^2 (Before Optimization):
{:.5f}".format(best_model.score(X_test, y_test)))

params = {
    'learning_rate': [0.01, 0.1, 1.0],
    'n_estimators': [100, 150, 200],
    'max_depth': [3, 4, 5]
}

clf = GridSearchCV(best_model, params)
clf.fit(X_train, y_train)

clf.best_params_

print("Model R^2 (After Optimization): {:.5f}".format(clf.score(X_test,
y_test)))
```

21. Comparing original values to values predicted by model

```
x_predict = list(best_model.predict(X_test))
predicted_df = {'predicted_values': x_predict, 'original_values': y_test}
#creating new dataframe
pd.DataFrame(predicted_df).head(53)
```

22. Predicting strength values for new input data

```
new_input = [[1,1,30,69,99,0,4]]
new_output = model.predict(new_input)
print(new_input, new_output)
```

23. Saving the model

```
import pickle
file = 'concrete_strength'
save = pickle.dump(best_model,open(file,'wb'))
```

References

[1] D. F. L. . L. Chen, " Stabilization treatment of soft subgrade soil by sewage sludge ash and cement," *Journal of Hazardous Materials,* 2009.

[2] L. DN, "Evaluation of structural properties of lime stabilised soils and aggregates," National Lime Association, 1999, pp. 1-89.

[3] Z. M. D. J. Solanki P, Resilient Modulus of clay subgrades stabilized with lime, class C fly ash, and cement kilndust for pavement design, Transp Res Rec J Transp Res Board 2186:101–110, 2010.

[4] S. R. G. Rajasekaran, "Strength characteristics of lime-treated marine clay," *Proceedings of the Institution of Civil Engineers,* vol. 4, no. 3, 2000.

[5] S. N. D. N. Little, "Recommended practice for stabilization of subgrade soils and base materials," 2009.

[6] K. K. H. N. A. Eisazadeh, ",Solid-state NMR and FTIR studies of lime stabilized montmorillonitic and lateritic clays," *Applied Clay Science,* 2012.

[7] R. G. J.L. Eades, "Reaction of hydrated lime with pure clay minerals in soil stabilization," *High. Res. Board Bull,* 1960.

[8] M. F. D. P. S. R. s. Ankit Singh Negi, "Soil stabilization using lime," *International Journal of Innovative Research in Science, Engineering and Technology,* vol. 2, no. 2.

[9] H. Q. K. S. J. Mallela, "Consideration of lime-stabilized layers in mechanistic-empirical," *The National Lime Association,* 2004.

[10] A. R. B. Association, "LIME STABILIZATION & LIME MODIFICATION," in *LIME-TREATED SOIL CONSTRUCTION MANUAL*, National Lime Association , 2003.

[11] T. DV, "Hydro-mechanical behaviour of lime-stabilised soils," UniversitÃ degli Studi di Cassino FacoltÃ di Ingegneria, 2006.

[12] H. B. B. R. Louafi B, "Improvement of geotechnical characteristics of clay soils using lime," *Advanced materials research, Trans Tech Publications, Zürich,* vol. 1105, p. 315–319, 2015.

[13] E. B. W. R. L. J. L. Jair A.Baldovino, "Effects of lime addition on geotechnical properties of sedimentary soil in Curitiba, Brazil," *Effects of lime addition on geotechnical properties of sedimentary soil in Curitiba, Brazil,* vol. 10, no. 1, 2018.

[14] C. G. O. A. A. F. &. M. S. B. Ali Akbar Firoozi, "Fundamentals of soil stabilization," *International Journal of Geo-Engineering,* 2017.

[15] A. H. Alavi, "Prediction of maximum dry density and optimum moisture content of stabilised soil using RBF neural networks," *The IES Journal Part A: Civil & Structural Engineering ,* 2009.

[16] P. S. &. A. K. S. Sarat Kumar Das, "Application of Artificial Intelligence to Maximum Dry Density and Unconfined Compressive Strength of Cement Stabilized Soil," *Geotechnical and Geological Engineering,* 2010.

[17] A. I. L. Ruhul Amin Mozumder, "Prediction of unconfined compressive strength of geopolymer stabilized clayey soil using Artificial Neural Network," *Computers and Geotechnics, Elsevier,* 2015.

[18] A. C. P. C. JoaquimTinoco, "Application of data mining techniques in the estimation of the uniaxial compressive strength of jet grouting columns over time," *Construction and Building Materials, Elsevier,* 2011.

[19] Z. B.-w. Shen Rong, "The research of regression model in machine learning field," *MATEC Web of Conferences ,* Jan, 2018.

[20] A. Gajurel, B. Chittoori, P. S. Mukherjee and M. Sadegh, "Database used in Machine Learning Methods to Map Stabilizer Effectiveness based on Common Soil Properties," Vols. Texas Data Repository, V3, 2021.

[21] F. G. Bell, "Lime stabilisation of clay soils," *Bull of the Inter. Asso. Of Engg. Geo.-Bulletin de ,* 1989.

[22] A. R. B. A. S. o. L. Stabilization, LIME-TREATED SOIL CONSTRUCTION MANUAL, National Lime Association, Jan, 2004.