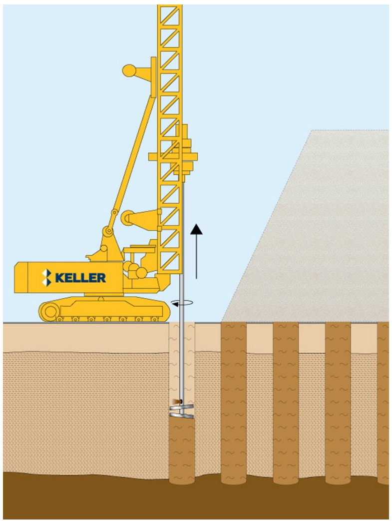
Norwegian University of Science and Technology, NTNU
Trondheim, Autumn 2022

Application of Data Mining Techniques to Soil Stabilization



Figur 0.1: Illustration of DDM machine

Markus Nikolai Berner

Supervisor: Yutao Pan, NTNU

Co-supervisor: Stefan Ritter, NGI

Co-supervisor: Dominik Gächter, Keller

TBM4500 - Civil and Environmental Engineering, Specialization Project

Abstract

Dry deep mixing is a soil stabilization technique commonly used in Norway. This is a complicated method, where many factors influence the strength increase. According to Terashi (1997), these factors are divided into four categories: Characteristics of stabilizing agent, Characteristics and conditions of soil, Mixing conditions, and Curing conditions.

Quality control is essential to ensure that the finished product has the mechanical properties as designed. During production, the execution parameters are recorded. This means that the contractor has large amounts of data where exciting patterns may exist related to the mechanical properties of the stabilized soil. Investigations within this topic are not very well-established.

With the help of machine learning, it is possible to find patterns in the data. This can be used to investigate what effect the various execution parameters have on the mechanical properties of a deep dry column. Multiple studies have been carried out in this field of study. However, none of the models used in these studies were based solely on the execution parameters. With the help of different algorithms, such as Artificial Neural Networks, Decision Trees, Support Vector Machines, and Multiple Regression, it was possible to create reliable models.

A dry deep mixing database has been created in this thesis, consisting of 30 columns from different projects. A preliminary analysis has been conducted. The correlation between BRN, Binder Dosage, and Cu-mean has been examined. The results indicate zero correlation to a weak negative correlation between the variables. It is challenging to interpret specific trends in the data, and it is not possible to give any particular advice for future DDM projects without further data on the soil conditions.

Contents

Abstract **i**

Acronyms **iv**

List of figures **v**

List of tables **vi**

1 Introduction **1**
1.1 Background 1
1.2 Scope 1

2 Literature Study **2**
2.1 Approach 2
2.2 Deep Soil Mixing 2
2.2.1 Mechanism of Stabilization of cement treated soil 3
2.2.2 Factors Affecting the Strength Increase 4
2.2.2.1 Characteristic of Stabilizing Agent 4
2.2.2.2 Characteristics and Conditions of Soil 5
2.2.2.3 Mixing Conditions 5
2.2.2.4 Curing Conditions 8
2.2.3 Quality Control 8
2.2.3.1 Execution parameters 9
2.2.3.2 Quality verification 11
2.2.3.3 Spatial Variability 14
2.3 Data Mining 16
2.3.1 Knowledge Discovery in Databases 16
2.3.2 Data-Mining Processes 16
2.3.2.1 The KDD Process 16
2.3.2.2 The SEMMA Process 17
2.3.2.3 The CRISP-DM Process 17
2.3.3 Data-Mining Methods 18
2.3.3.1 Classification 19
2.3.3.2 Regression 19
2.3.3.3 Clustering 19
2.3.4 Data-Mining Algorithms 21
2.3.4.1 Artificial Neural Networks 21
2.3.4.2 Decision Trees 23
2.3.4.3 Support Vector Machines 24
2.3.4.4 Multiple Regression 24
2.3.5 Performance of the models 25
2.3.5.1 The Squared Correlation Coefficient, R^2 25
2.4 Comparison of results from articles in litterateur review 26

3 Presentation of Dataset **29**

4 Results **30**
4.1 Scatter Plots and Correlation 30

5 Summary and Conclusion **34**
5.1 Recommendation for Further Work 35

Bibliography **36**

Acronyms

DDM	Deep Dry Mixing
OPC	Ordinary Portland Cement
UCS	Unconfined Compressive Strength
MDD	Maximum Dry Density
E	Young's modulus
BRN	Blade Rotation Number
CPT	Cone Penetration Test
PIRT	Push-In Resistance Test
KSP	Kalk-Pelar-Sondering
FOPS	Förinstallerad Omvänd Pelar Sondering
PORT	Pull-Out Resistance Test
DM	Data Mining
KDD	Knowledge Discovery in Databases
SEMMA	Sample, Explore, Modify, Model and Assess
CRISP-DM	CRoss-Industri Standard Process for Data Mining
ANN	Artificial Neural Networks
DT	Decision Trees
RF	Random Forest
GB	Gradient Boosting
SVM	Support Vector Machines
MR	Multiple Regression

List of figures

0.1 Illustration of DDM machine 1
2.1 Example of a production log 10
2.2 Summary of the data from the production log 11
2.3 Illustrations of probes used for in situ testing, according to Larsson (2005). Section A-A shows the different cross-sections of the blades. 13
2.4 Example of KSP log 13
2.5 The KDD process according to Fayyad et al. (1996a) 16
2.6 The CRISP-DM process according to Olson and Delen (2008) 18
2.7 Presentation of the data-mining method of Classification 20
2.8 Presentation of the data-mining method of Regression 20
2.9 Presentation of the data-mining method of Clustering 21
2.10 ANN 22
4.1 Curing days Vs Cu-mean 30
4.2 Binder dosage Vs Cu-mean 30
4.3 BRN Vs Cu-mean 31
4.4 BRN Vs Binder dosage 31
4.5 Quality control for Column nr. 26 33

List of tables

2.1 Factors affecting the strength increase according to Terashi (1997) 4
2.2 Summary of Larsson's opinions on how much influence the various factors have on the mixing process, obtained from Larsson (2005). The different symbols under strength variability mean:
+++ Significant and major influence
++ Significant influence
+ Divergent results
- No or weak influence 7
2.3 Summary of articles used to study data mining techniques applied to geotechnical engineering 26
2.4 Visualisation of the different Data Mining techniques used and the performance of the model. The R^2 value given applies to the model that gave the best result. Under performance of the model, the acronyms are given as:
UCS = Unconfined Compressive Strength
E = Young's modulus (Stiffness)
MDD = Maximum Dry Density 26
2.6 Overview of Dataset, Input, and Output 27
2.5 Overview of Dataset, Input, and Output 28
3.1 Summarization of Data provided from Keller Geoteknikk 29
4.1 Correlation matrix between CU-mean, BRN, and Binder dosage 32

1 Introduction

1.1 Background

Deep dry mixing (DDM) is a soil stabilization technique that is common in Scandinavia. This method uses cement and lime to stabilize weak soils. The contractors use sophisticated machinery that collects data from the soil stabilization process. The data is often related to the rotations per minute, penetration and retrieval rate of the drill string, the binder dosage and the air pressure. This data is used mainly for quality control, but it is possible that there are some connections in this data that have not yet been discovered. Here it is interesting whether these discoveries can say something about the mechanical properties of the stabilized soil.

1.2 Scope

The tasks for this project are divided into five sub goals:

- Study data mining techniques applied to geotechnical engineering.
- Acquire and systemize real-life DDM data to obtain a DDM database.
- Apply relevant data mining techniques to the derived DDM database.
- Test if reliable correlations in the DDM database can be detected.
- Translate the obtained findings into practical guidance informing future DDM projects.

This project thesis is the start of the work to be carried out in the Master's thesis in the spring 2023. Since a large part of the project assignment deals with literature study, the first bullet point from the list above is therefore the most important. It will further be carried out a preliminary analysis which deals with the next four tasks. The complete analysis will be continued in the master's thesis.

2 Literature Study

2.1 Approach

A deep dive into the theory regarding Data Mining techniques and soil stabilization is required to write a thesis about Data Mining techniques and soil stabilization. Therefore a literature study was conducted to obtain a good enough theoretical background within the field of study.

In this thesis, it has been looked at the research gap on how the execution parameters affect the mechanical properties of the stabilized soil. A lot of relevant literature was provided by the supervisors, but additional information and articles were found using Google Scholar and Oria. When utilizing these search engines, the keywords "Soil stabilization", "Deep Soil Mixing", "Data Mining", and "Machine learning" were applied.

2.2 Deep Soil Mixing

Deep soil mixing is a ground improvement technique that is used in many countries all over the world. Increased mechanical and physical properties characterize the stabilized soil. This means that the soil will have a higher strength, a lower permeability, and give less settlements (Keller (2022)). Deep soil mixing is divided into two categories: *Dry deep mixing* and *Wet deep mixing*. In Norway and other Nordic countries, it is familiar with dry deep mixing. This method will transport the binder to the soil with compressed air as the medium. This is because these countries' soil is often loose and has a high water content. Thus, it makes sense not to add more water. This means that it will be possible to stabilize very sensitive soils.

Nevertheless, compressed air can present some difficulties. According to Larsson (2014), the dispersion process is complicated, which is the part where the binder is distributed into the soil. According to Keller (2022), the typical diameter for a dry deep mixing column ranges from 0,6 to 1m, with a maximum depth of 25 m. On the other hand, wet deep mixing differs from dry deep mixing in that the binding agent is mixed with water to form a slurry. With the help of special mixing tools, the slurry is injected into the ground. Wet deep mixing will have the same maximum depth as dry deep mixing, but the wet method can make columns with a diameter of up to 2.4 m (Keller (2022)). It is possible to reinforce the columns to obtain an even greater bending resistance. This will be particularly relevant if excavation is to be carried out up to the pillars.

2.2.1 Mechanism of Stabilization of cement treated soil

The most common type of cement is the Ordinary Portland Cement (OPC), with tricalcium and dicalcium silicates (C_3S and C_2S), tricalcium aluminate C_3A and tetracalsium alumino-ferrite C_4AF as its main components (Huawen (2009)). In cement chemistry, it is common to have notations for the different components. According to MacLaren and White (2003) is:

- C = Calcium oxide (lime), CaO
- S = Silicone dioxide (silica), SiO_2
- A = Aluminium oxide (alumina), Al_2O_3
- F = Iron(III) oxide, Fe_2O_3

The tricalcium and dicalcium silicates are very reactive to water, and they will produce colloidal hydrated products. Of the two silicates, tricalcium is the best component because it is more reactive, which means it uses less time to set and for the hydration process. The soil's stabilization can be divided into two main chemical reactions: The primary hydration and the pozzolanic reaction. The primary hydration is shown in 2.1 and 2.2. 2.1 shows what happens when the tricalcium reacts with water. The result of this reaction is the formation of CSH, which is a hydrated gel that leads to a gain in strength in the soil. This reaction is categorized as a short-term hardening that typically takes 7-28 days. In the reaction between $C_3S + H_2O$ there will also lead to the formation of $Ca(OH)_2$, as seen in 2.1. This will lead to an increase in concentrations of $Ca^{2+} + 2(OH)^-$ as seen in 2.2, which is the result of the hydrolysis of the lime.

$$C_3S + H_2O \rightarrow C_3S_2H_x + Ca(OH)_2 \tag{2.1}$$

$$Ca(OH)_2 \rightarrow Ca^{2+} + 2(OH)^- \tag{2.2}$$

The next step of the hardening of the cement is the secondary pozzolanic reaction. This is categorized as a long-term hardening that happens after 28 days. Here the concentration of $(OH)^-$ is central. When the pore water has enough $(OH)^-$ ions will, this facilitates a number of reactions. This is shown in reaction 2.3 and 2.4. These reactions occur due to the $(OH)^-$ ions will react with the clay, which leads to a dissolution of silica 2.3 and alumina 2.4. In both these reactions, the end product will be secondary cementatious products. With time these products

(I) Characteristic of stabilizing agent	1. Type of hardening agent 2. Quality 3. Mixing water and additives
(II) Characteristics and conditions of soil	1. Physical, chemical and mineralogical properties of soil 2. Organic content 3. pH of pore water 4. Water content
(III) Mixing conditions	1. Degree of mixing 2. Timing of mixing / re-mixing 3. Quantity of hardening agent
(IV) Curing conditions	1. Temperatur 2. Curing time 3. Humidity 4. Wetting and drying / freezing and thawing, etc.

Table 2.1: Factors affecting the strength increase according to Terashi (1997)

will harden, which leads to a further increase in strength in the stabilized soil.

$$Ca^{2+} + 2(OH)^{-} + SiO_2 \rightarrow C - S - H \tag{2.3}$$

$$Ca^{2+} + 2(OH)^{-} + Al_2O_3 \rightarrow C - A - H \tag{2.4}$$

2.2.2 Factors Affecting the Strength Increase

Soil stabilization is a complicated process where many factors influence the final strength increase. Terashi (1997) has divided these factors into four main categories: *(I) Characteristic of stabilizing agent, (II) Characteristics and conditions of soil, (III) Mixing conditions, and (IV) Curing conditions*, all of which have several subcategories. The different affecting factors are visualized in 2.1.

2.2.2.1 Characteristic of Stabilizing Agent

According to Larsson (2005), the strength of the stabilized soil will be increased with a higher quantity of binder. However, it must be said that there is not a linear relationship between improved strength and added binder. A further addition will not lead to a higher strength at a given percentage of the added binder. Huawen (2009) refers to various studies where it has been found that the best increase in strength of the soil was carried out with different cement contents. This is because the strength of the soil is not only determined by the amount of binder but is rather a relationship between soil, cement, and water dosage. According to Zhang et al.

(2018), a higher dosage of clay will decrease strength. The clay dosage significantly influences the strength with a higher water/cement (w/c) ratio. The best conditions for soil stabilization are with the lowest clay dosage and w/c ration. Furthermore, it has been shown that different types of binder can lead to different results in compressive strength (Huawen (2009)).

2.2.2.2 Characteristics and Conditions of Soil

Due to the large differences in mechanical properties in different soils, there are also naturally large differences in the properties after the soil is stabilized. This is because the chemical reactions that occur between the soils and the binders vary between the various types of soil. Huawen (2009) refers to several studies that have been done within this field. Based on this Huawen (2009) found that the grain size distribution greatly influenced the unconfined compressive strength (UCS). According to Boutouil and Levacher (2005), the initial water content has a great influencing force on the UCS. In this context, a lower initial water content will mean a higher UCS. The organic content also has a great influence on the UCS. With a greater content of organic material, this will lead to a smaller total compressive strength (Pradeep and Vinu (2015)). When it comes to the pH in the pore water, this will affect the UCS of the stabilized soil. As said in 2.2.1, the concentration of $(OH)^-$ is central for the secondary pozzolanic reaction. This means that a higher pH value will facilitate the chemical reactions 2.3 and 2.4, which leads to an increase in strength.

2.2.2.3 Mixing Conditions

The mixing process is a complicated process, which can be summed up as the spread of binder in the soil. Here it is desired to have a good enough mixing of the soil so that the chemical reactions between the binder and the soil can occur. According Larsson (2005), many factors may influence the mixing process:

- Mixing energy
- The binder and the amount of binder
- Mixing tool geometry
- The rheology of the soil
- Delivery pressure and the amount of air
- Compaction energy

The mixing energy is something that is normally not measured, but it is possible to calculate this energy based on some execution parameters. The mixing energy or Blade Rotation Number (BRN) is a combination of the total number of blades and the retrieval rate [mm/rev] (Larsson (2005)), as seen in 2.5:

$$BRN = \sum Nr.blades * \frac{1}{Retrieval\,rate} \tag{2.5}$$

According to Larsson (2005), many studies state that the number of blades, in combination with the retrieval rate, have a major role in the strength and the coefficient of variation of the stabilized soil. This means that if the BRN is increased, the strength will also increase. While on the other side, the coefficient of variation will decrease. Based on this, it looks like increasing the BRM will be the natural solution to increased strength. Here, on the other hand, Larsson (2005) believes this solution is not necessarily the best because the mixing process is very complex, with many factors influencing the result. Therefore it is simpler to adjust the binder and the binder quantity to adjust the strength.

The mixing tool design is very important for the uniformity of the dry deep mixing column. It needs to be designed to facilitate even distribution of the binder. This gives the columns a uniform cross-section throughout the entire depth, leading to uniform mechanical properties.

The rheological properties of the soil are other factors that influence the mixing process. This applies particularly to fine soils, like clay, since they are very cohesive. These soils can give challenges to both the mixing and also the monitoring of the process. Because of this, it is common to intensify the mixing process to be able to spread the lime and cement into the cohesive soil.

Another important factor that affects the mixing process, according to Larsson (2005), is the delivery pressure and the amount of air. Larsson (2005) says that the air pressure should be a middle ground between low and high air pressure. This is because the air pressure needs to be high enough so that the air can make a path up to the surface, which prevents the air pressure from building up under the ground. On the other side, the initial air pressure mustn't be too big because this can lead to pneumatic fracturing outside the column periphery (Larsson (2005)).

The compaction energy is another important factor that influences the mixing conditions. In Larsson (2005), it discusses how early the dry deep mixing columns should be loaded after their construction. Applying loads to the columns at an early stage will it be possible to improve the strength. This is because, when the columns are compressed, the soil will become denser by the air being forced out, which results in a reduction of the volume of voids. An early compression will also help to consolidate the soil, which is that the water is forced out of the voids. An early compression may lead to even better results in Scandinavia and other countries that use air to transport the binder to the soil. This is because the air used to transport the binder can be trapped under the ground.

Factor	Strength variability
Mixing tool geometry related to incorporation and spreading of binder	+++
The rheology of the soil	+++
Compaction and consolidation	+++
Retrieval rate	++
Number of blades	++
Binder content	++
Amount of air	++
Mixing tool geometry	+
Type of binder	+
Rotation speed	-
Air pressure	-

Table 2.2: Summary of Larsson's opinions on how much influence the various factors have on the mixing process, obtained from Larsson (2005). The different symbols under strength variability mean:
+++ Significant and major influence
++ Significant influence
+ Divergent results
- No or weak influence

The factors that influence the mixing process are summarized in 2.2. This table is taken from Larsson (2005) and is based on the author's opinions in combination with litterateur reviews. Here one can see that three factors especially impose a significant and major influence on the mixing process. These are The compaction and consolidation, the rheology of the soil, and the mixing tool geometry related to the incorporation and spreading of the binder. Based on 2.2, it may seem unclear that the mixing tool geometry is mentioned twice, where the ordinary mixing tool factor has got the strength variability (+), while the mixing tool geometry related to incorporation and spreading of the binder has got (+++). This is because no thorough studies have been carried out regarding this factor, and those that have been made are more sales-oriented than scientific (Larsson (2005)).

Furthermore, it is interesting to see that it is not the air pressure but rather the amount of air that influences the strength variability. The rotation speed is also a factor with no or weak influence.

This is because many studies have not been carried out on the influence of rpm when it comes to strength magnitude. Usually, you want as high a rotation speed as possible since this will speed up the installation process. In Scandinavia, it is common to have rotation speeds from 150-200 rpm.

The retrieval rate and the number of blades are the execution parameters that are used to calculate the blade rotation number. An increase in BRM is known to increase strength. It is also shown in 2.2 that it is rather the binder content instead of the type of binder that influences the strength the most.

2.2.2.4 Curing Conditions

The curing conditions are the last category that affects the strength in a DDM column according to Terashi (1997). Huawen (2009) has looked into this category and found that the temperature, curing time, humidity, and curing stresses affect the strength development. The curing temperature is an important factor. It has been shown that a higher temperature at curing will lead to a higher strength. Here, an almost linear relationship exists between the curing temperature raging between 0-30 degrees Celsius and the UCS.

The curing time plays an important role in the increase in UCS. The longer time the column is given to harden, the greater increase in UCS is to be expected. This is because the longer the column has for curing, this will lead to more time for the pozzolanic reactions to happen (Ghee (2006)).

The effect of curing stress varies depending on the ground conditions. Huawen (2009) refers here to the fact that UCS will increase with the confining stress if the DDM column is constructed under drained conditions. On the other hand, there will be no increase in UCS if the column is cast under undrained conditions.

Finally, there is the humidity. Here it is also shown that an increase in humidity will also lead to an increase in UCS (Huawen (2009)).

2.2.3 Quality Control

Quality control is important when executing DDM works. DDM columns differ from other piles, such as steel and concrete, and the strength of the finished product is a combination of several factors. This means that there are strict requirements for the correct execution of quality control so that the contractor can verify that the product provides the design strength. According to

Larsson (2005), quality control can be divided into:

- Laboratory tests
- Field tests on the test columns
- Quality control during execution
- Quality verification after execution and follow-up measurements

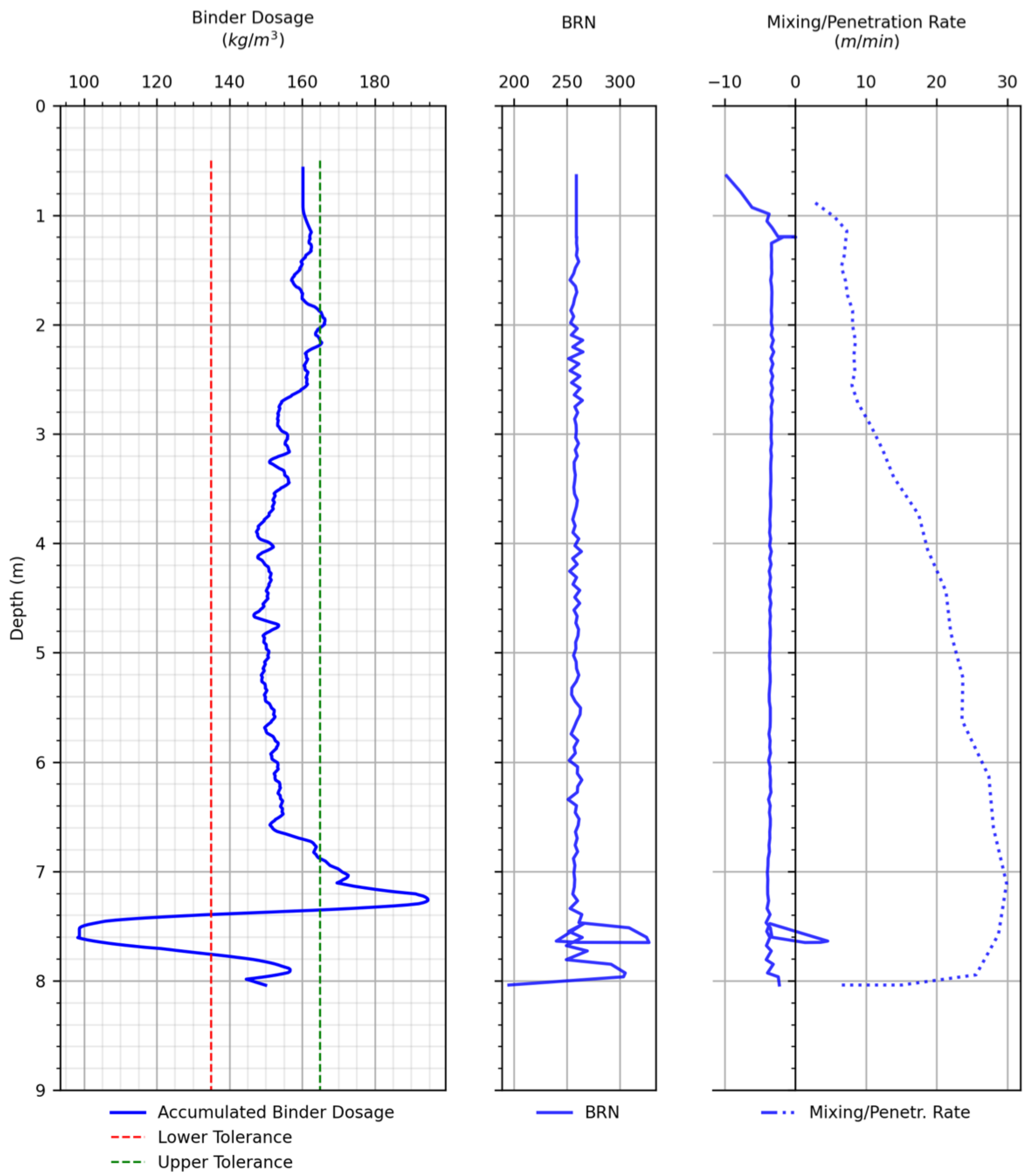
It is common to carry out laboratory tests on the soil before production occurs. The problem is that the samples that are tested in the laboratory do not necessarily reflect the characteristics of the soil to be stabilized. Nevertheless, this can lay the foundation for the parameters that can be used in a given project. When the preliminary execution parameters are determined, a test column is created on the site. In this test column, several field tests are carried out. Based on the results from the laboratory and the test column, the final mix design is completed.

2.2.3.1 Execution parameters

During the installation of the column, several parameters are monitored and recorded. These are often visualized in a chart-log, as seen in 2.1 and in the accompanying table 2.2. Typical execution parameters according to Keller (2022) are:

- Element identification: Name of column and project number.
- Mixing tool details: Name of mixing tool, the diameter of the column, and the total number of blades.
- Mixing depth: This shows the depth of the column, which is the same as the stabilized depth. It is also common to give information about the surface elevation or reference height and the height of the top of the column.
- Mixing time: The total time used to carry out the work.
- Binder specification: Type of binder that has been used for the work, for instance, OPC, fly ash, ex.
- Binder dosage and air pressure.
- Total volume of binder used.
- Mixing tool velocities (The penetration rate) and rpm during penetration.

- Withdrawal and torque of the shaft.
- Blade rotation number: As presented in 2.2.2.3.



Figur 2.1: Example of a production log

As seen in 2.1, the various execution parameters are visualized. These graphs show how the binder dosage, BRN, and mixing/penetration rate vary with depth. Here, the BRN, binder dosage, and mixing rate are almost constant with depth. The penetration rate, on the other hand, increases with depth.

Equipment:	822	**Start Date:**	
Mixing Tool:	pb3	**Start Time:**	14:16:52
Stabilized Length:	7.48 m	**Completion Date:**	
Column Diameter:	0.60 m	**Completion Time:**	14:20:02
Surface Elevation:	23.25 m	**Number of Blades:**	6
Top of Column:	0.56 m	**Max. Drill Depth**	8.04 m
Total Binder Weight:	326 kg	**Avg. Binder Dosage:**	154.12 kg/m³
Avg. Lift Coeff.:	-22.12 mm/rev	**Avg. Pull Rate:**	-3.11 m/min
Avg. RPM:	141	**Avg. BRN:**	260
Tank Pressure at start:	7.81 bar		

Figur 2.2: Summary of the data from the production log

In 2.2, the data from the production log in 2.1 is summarized. These values are typically given as average numbers. Furthermore, examples of values for the other execution parameters are also provided in 2.2.

2.2.3.2 Quality verification

The stabilized soil needs to be tested to verify that the DDM columns have the correct mechanical properties. This can either be done in the laboratory or in situ. The results obtained in the laboratory will usually give a higher strength than in situ testing, with the same amount of binder. This is because the mixing done in the laboratory will be more consistent than in the field. However, the results from the field mixing will be influenced by overburden pressure and higher curing temperature, which will give a higher strength. The difference in strength between the in situ and laboratory results is proven to be around 50 percent (Santos Barros (2019)).

A test sample from the field must be obtained to test the soil in the laboratory. This can either be a core sample of the stabilized soil or in situ soil mixed with the binder in the laboratory. Typical tests that are done in the laboratory are:

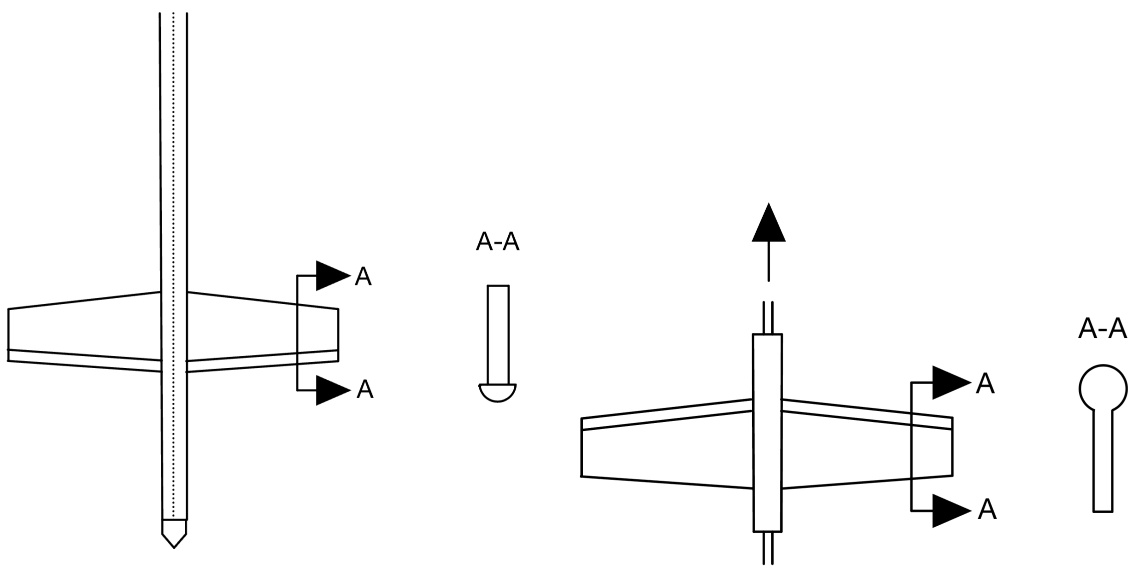
- Unconfined compression test (UCS)
- Triaxial test
- Oedometer test

The laboratory test can be useful in the early stages of the design. This is done to test the soil with various types of binders and to find the correct content of the binder. It is also recommended to perform in situ tests of the final product. There are different types of in situ tests that can be carried out:

CPT or Cone Penetration Test is a method that can be used to determine the in situ mechanical properties of the stabilized soil. At the tip of the probe, it will be monitored and recorded the penetration resistance, sleeve friction, and the pore water pressure (Larsson (2005)). Since the stabilized soil has high strength, this can cause problems with the penetration of the probe in the column. Therefore it is possible to drag the probe through the column, which is similar to the procedure in KPS and FOPS.

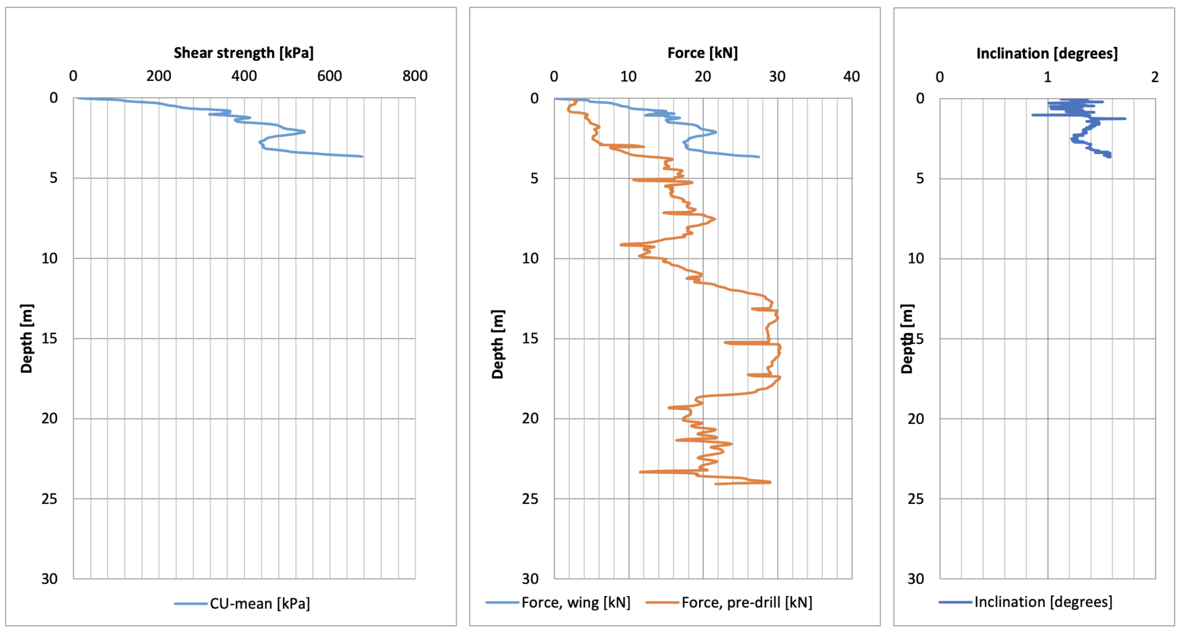
KPS is a test method used for DDM columns. KPS is a Swedish word and stands for "Kalk-Pelar-Sondering." The method is also known under the name PIRT, which is an acronym for push-in resistance test. From this test, it is possible to assess the strength of the stabilized soil. This is done by pressing a probe down through the column at a constant speed of 20 mm/s. When this is done, the penetration resistance is recorded. The design of the probe can be described as two airplane wings, as seen in 2.3a, which should be so wide that they are precisely 100 mm smaller than the diameter of the column. Should problems occur with pressing the probe into the column, it is possible to pre-drill a center hole. On the other hand, if this is not enough to push the probe through the column, it is possible to use dynamic impact. The PIRT method will be advantageous over CPT since CPT only will give a result in a point. When the CPT probe is pressed down, it tends to follow the weakest path.

FOPS is a method that is similar to KPS. This is also a Swedish word and stands for "Förinstallerad Omvänd Pelar Sondering," which is translated into English and means "Pull-out resistance test" (PORT). The method separates from KPS because the probe will be installed during the mixing process. The probe used for FOPS also has a similar shape as the one used for KPS, but instead of having rounded edges on the bottom of the probe wings, these rounded edges are on top (see 2.3b). When the DDM column has set, the probe will be pulled from the bottom and up using a steel cable. According to Larsson (2005), FOPS testing tends to give less reliable results since the installed steel wire will disturb the mixing process. However, the results will be trustworthy if the probe is installed after manufacturing the column.



(a) Illustration of probe used for KPS **(b)** Illustration of probe used for FOPS

Figur 2.3: Illustrations of probes used for in situ testing, according to Larsson (2005). Section A-A shows the different cross-sections of the blades.



Figur 2.4: Example of KSP log

The graph to the left shows the shear resistance, CU-mean. The middle graph shows the resistance force for both the pre-drilling and the wing. The graph to the right indicates the inclination of the probe during the penetration. According to Timoney and McCabe (2017), the shear strength is given by the equation 2.6.

$$C_u = \left(\frac{1}{N}\right)\left(\frac{P}{A}\right) \tag{2.6}$$

In 2.6, N is a bearing capacity factor, usually 8-10. P is the penetration force seen in the middle graph in 2.4. The last factor is A, which is the cross-sectional area of the probe. This area includes both the wings and the cone of the probe.

2.2.3.3 Spatial Variability

As said, the stabilization process is complicated and contains many different factors and steps. When this process is completed, it is desired to measure the mechanical properties of the stabilized column. The problem is that these properties are influenced by both execution and mixing. During the mixing process, the most uniform spread of the binder and mixing of the soil is desired. This will result in good mixing quality and low variability. There are not any specific guidelines that are provided to ensure good enough variability. Larsson (2005) says that the variability is handled by adjusting the rotation speed, the retrieval rate, and the penetration rate.

The stabilized soil will be in-homogeneous since the natural soil is this in the first place. This means that the stabilized soil will have properties that are anisotropic and non-elastic (Larsson (2005)), which again means that the stabilized soil will have complicated properties regarding spatial strength. The spatial variability is that the soil properties will vary across the mass of the soil. In many cases, it is common to assume average values when performing calculations. This can, in some cases, lead to overestimating or underestimating the actual safety factor. In contrast, different parameters are assigned a given statistical distribution when using spatial variability analysis. They will also be given correlation lengths for the parameters in both the x- and y direction. From the statistical distribution and the correlation length, many random values are calculated and then assigned to the soil mass (Rocscience (2022)). This means that the parameters will vary through the mass of the earth, which will be more like it is in reality.

According to YUTAO (2016), cement-stabilized soil is a typical material with spatial variability. During the mixing process, several things can happen that will affect the global strength. If this process is not done well enough, it can lead to the formation of lumps inside the stabilized column. These lumps will have considerably less strength than the rest of the column, which leads to a reduction of the global strength. This can also happen if the spreading of the binder is not done correctly. This places demands on the operator of the DDM machine, in that they have to monitor the supply of binder continuously. Since a sudden drop in the added binder can lead to a reduction in the diameter of the column or other structural weaknesses.

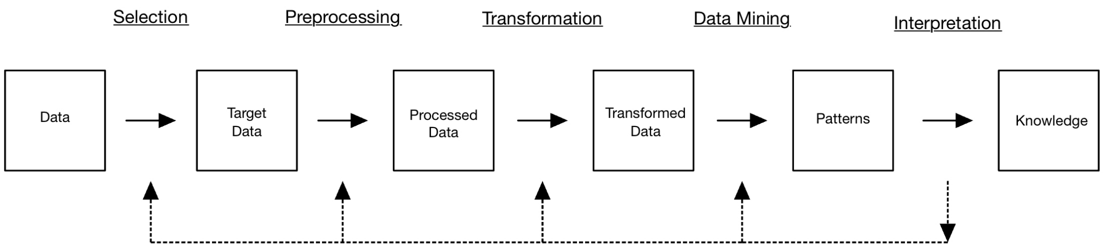
2.3 Data Mining

2.3.1 Knowledge Discovery in Databases

Today we live in a world where digital information is all around us. We get constant information from our cell phones, smartwatches, and other devices, which are stored and accumulated in databases. This also applies to science and businesses with access to large amounts of data. With the use of knowledge discovery in databases (KDD), it is possible to find useful knowledge from all this stored data. According to Fayyad et al. (1996b), KDD is "the overall process of discovering useful knowledge from data." Data mining (DM) is a central step of the KDD process. This is a process where the task is to interpret various patterns from the data using different methods and algorithms (Fayyad et al. (1996a)).

2.3.2 Data-Mining Processes

2.3.2.1 The KDD Process



Figur 2.5: The KDD process according to Fayyad et al. (1996a)

According to Mariscal et al. (2010), a KDD process is a model that contains several predetermined steps that the user needs to follow to perform a DM project. Mariscal et al. (2010) also describes the KDD process as proposed from Fayyad et al. (1996b) as the base model. This model is divided into multiple steps. The first one is called selection. Here the user needs to define which data from the database will be relevant to reach the goals for the KDD process. This places demands on the level of prior knowledge of the user so that the correct data is selected and is referred to as the target data. This data will go through pre-processing step, which intends to clean up the data. Basic operations are carried out here to remove outliers and to find methods to compensate for missing or unknown data. The next step is to prepare the data for the DM stage, which is done through a transformation process. When this is completed, the desired DM method is selected. The method must be chosen carefully to achieve the goals set at the start of the KDD process. Examples of different methods are Classification, regression, and clustering. In addition, a DM algorithm must be selected. The next step is DM, which looks for patterns in the transformed data. One of the most important parts of the DM step is to evaluate whether the patterns discovered are useful. Interpretation of these patterns may be turned into useful information for the user. On the other hand, if the patterns did not give the user the knowledge they wanted, the process may be repeated by going through any of the mentioned steps.

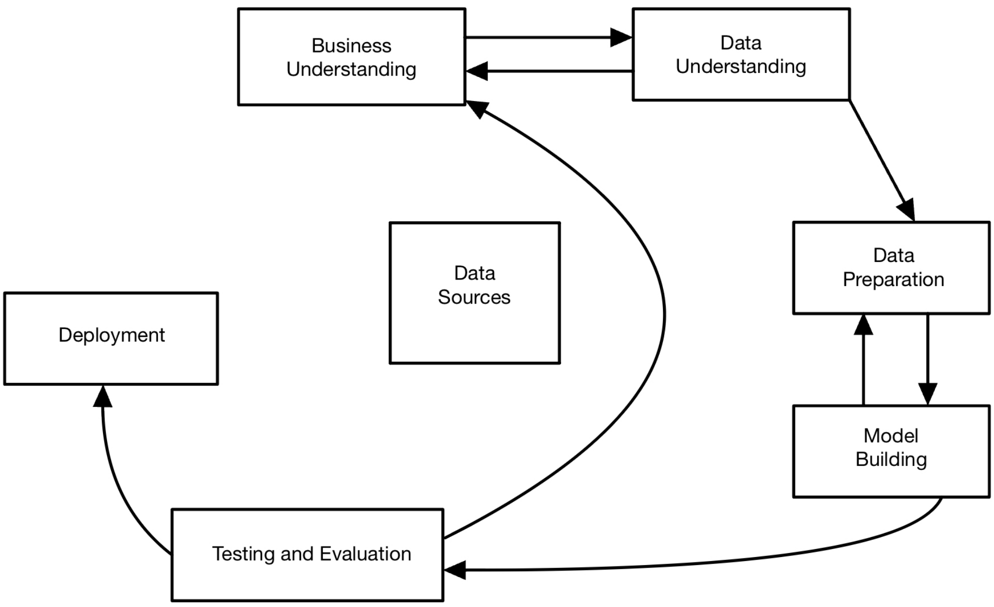
2.3.2.2 The SEMMA Process

SEMMA is another process that is used in a DM project. This model was developed by the SAS Institute and is an acronym that stands for *sample*, *explore*, *modify*, *model* and *assess* (Olson and Delen (2008)). In the first step, the data is selected. The choice of data has been made with care and must contain enough data to make the connections you want to find feasible. On the other hand, fewer data make it easier to manipulate the data set quickly. The total computation time will be reduced by reducing a very large data set to a smaller one. The next step is to explore. Here the user will go through a process that involves whether it is possible to learn something based on the data. This is done by looking for anomalies and trends that had not been foreseen in advance. The next step will start when the user has a better understanding of the data. In this step, the data will be modified by altering the variables. For example, removing various outliers that may interfere with finding patterns in the data may be applicable. It must also be considered if the number of variables is sufficient or whether more variables should be included or some of them removed. After the data is modified, it can be used in the model step. Various DM techniques can be used to find the desired output, which is found by searching through the data set. The discoveries made by data mining are evaluated in the final step. Here it is important to determine whether what was produced can be trusted.

2.3.2.3 The CRISP-DM Process

CRISP-DM is an abbreviation for **CR**oss-**I**ndustri **S**tandard **P**rocess for **D**ata **M**ining (Azevedo and Santos (2008)). The process is divided into six phases: *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation* and *Deployment* (Olson and Delen (2008)) as seen i 2.6. According to Piatetsky (2014), who conducted a poll in 2014, found out that over 40 percent of the participants used CRISP-DM when they conducted DM. SEMMA got around 9 percent in the same poll, while the KDD process got almost the same result. This means that the CRISP-DM process is one of the most popular.

As said, the process first starts with a Business Understanding. This means the user needs to figure out what they want to explore and why this new information is valuable. When a project plan is made, and the user has determined the main objectives of the DM, the process continues into Data Understanding. This is about acquiring relevant data. From there, you want to see if it is possible to find any preliminary patterns. Another important task in this phase is ensuring that the data is of high enough quality. When the data is selected, the process moves to the next phase, Data Preparation. This phase is similar to those already described in the KDD- and SEMMA processes and is all about cleaning the data so that it can be used for DM, which is the modeling phase. This is the phase where that data will assessed. Many different types of methods and algorithms can be used to analyze the relationships in the data. The phase aims for the user to acquire knowledge from these relationships. The gained knowledge is then compared with the user's hypotheses set in the first phases. The user then needs to figure out what these results mean and how the results are going to be used. In the final phase, the knowledge that has been obtained from this whole process is deployed.



Figur 2.6: The CRISP-DM process according to Olson and Delen (2008)

2.3.3 Data-Mining Methods

Many different DM methods are common to use. These methods are divided into two categories: Descriptive and predictive (Fayyad et al. (1996a)). The difference between these two mining types is that descriptive mining analyzes what has already happened from the available data. On the other hand is the predictive model, which uses the already existing data to predict what will happen in the future. It must be said that the border between the two models is not set in stone, and it is possible to have DM methods with both predictive and descriptive behavior.

2.3.3.1 Classification

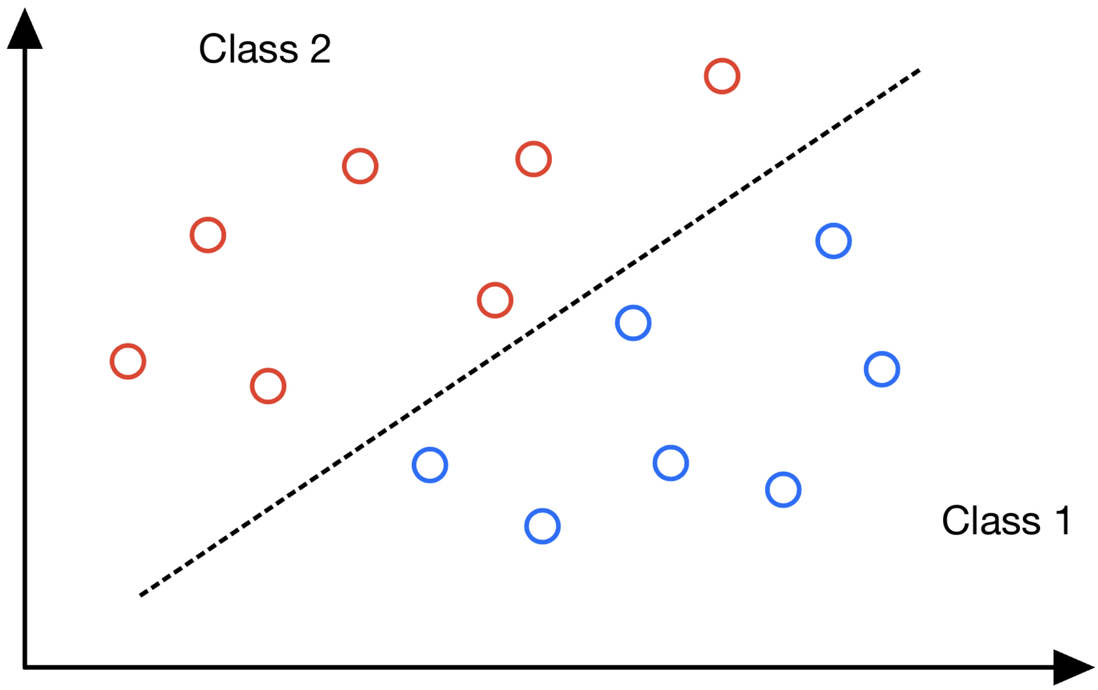
According to Olson and Delen (2008), Classification are functions that can learn how to sort data into different predefined sets of classes. The method will be trained through a learning set and presented to the different classes. If this training process is done correctly, it can classify an unclassified data set. An example of classification is shown in 2.7. The illustration presents how the data set is separated into two classes and how the classes are separated by using a linear boundary. Artificial neural networks and decision trees are examples of mathematical algorithms used to classify data.

2.3.3.2 Regression

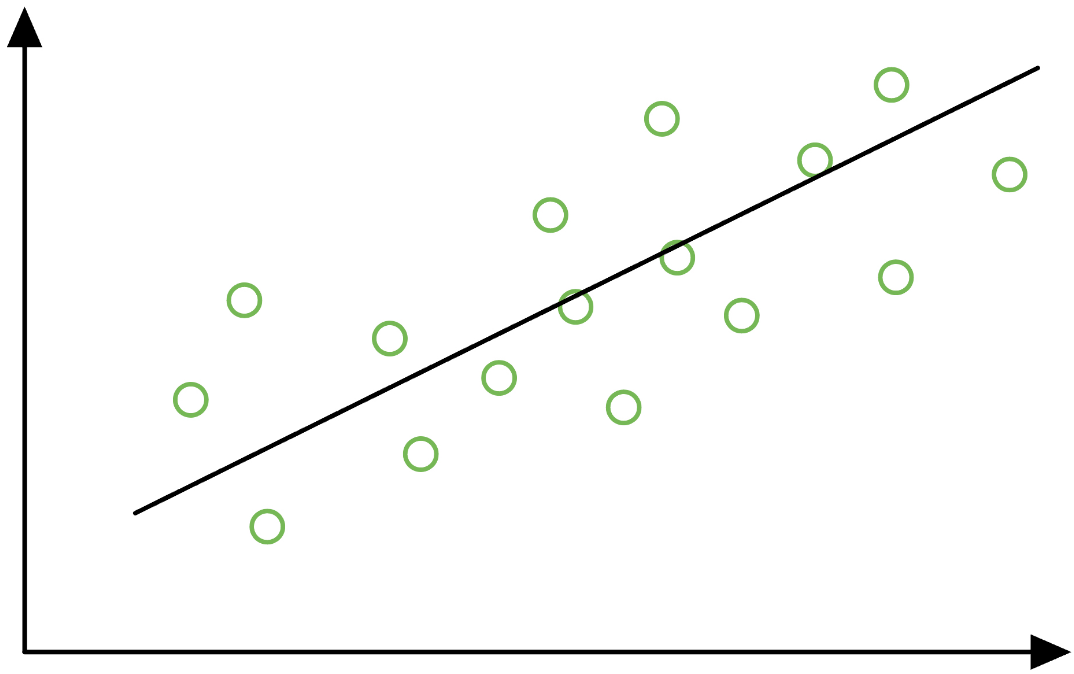
Another common DM method is Regression. This method will look at the relationship between the data and, based on this, produces a function that passes through the points in a matter that gives the lowest error. In other words, a predictive method that tries to say something about what will happen in the future. Regression analysis is used in many fields, like statistics and finance. A simple linear regression case is illustrated in 2.8. Based on the black line drawn through the data set, this suggests that there is an increasing trend.

2.3.3.3 Clustering

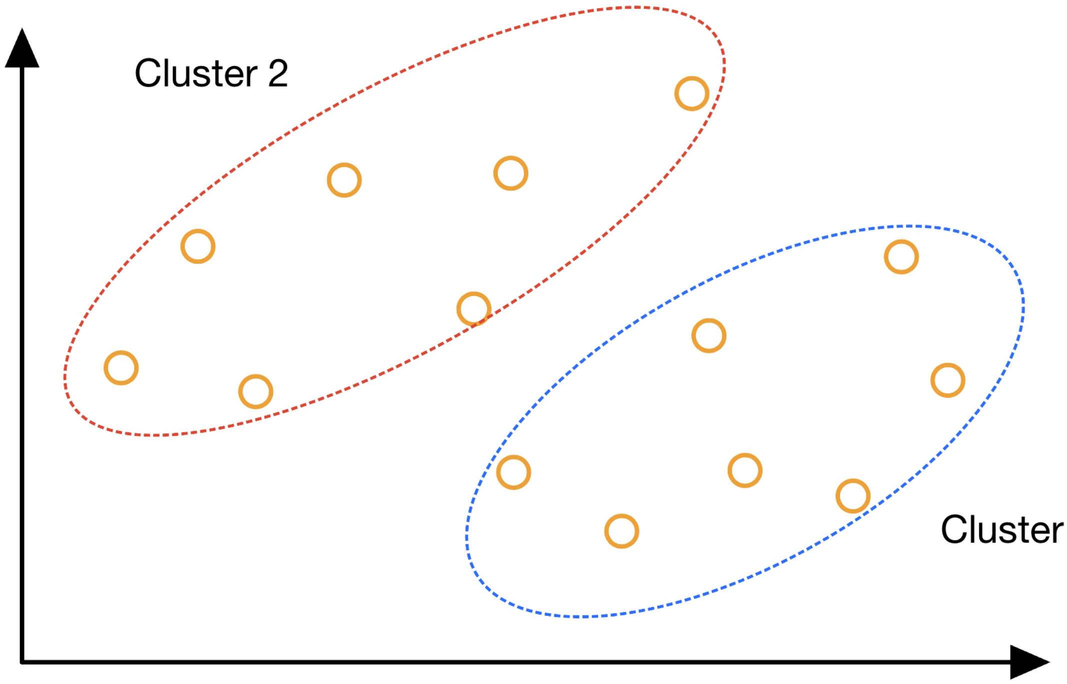
Clustering is another method that is used in DM. This method is unlike regression and classification because it is descriptive and not predictive. Nevertheless, clustering can be reminiscent of classification. The reason is that clustering also sorts data into separate groups, but the main difference is that these groups/classes are not predefined. In other words, the data will be analyzed and classified into separate clusters using special techniques. This will be done without training the model with a learning set (Olson and Delen (2008)). Unlike 2.7, where the data is separated into its original assigned class (Red and blue), are the data in 2.9 unlabeled (orange). The data will be put into clusters through the clustering process, as seen in the figure under 2.9.



Figur 2.7: Presentation of the data-mining method of Classification



Figur 2.8: Presentation of the data-mining method of Regression



Figur 2.9: Presentation of the data-mining method of Clustering

2.3.4 Data-Mining Algorithms

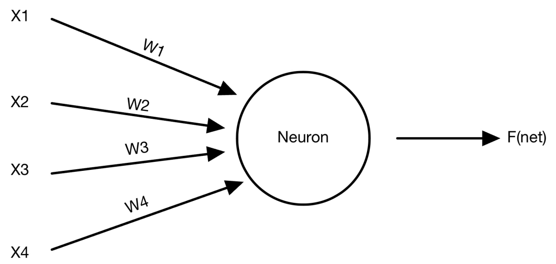
To be able to do a DM analysis, the methods (Clustering, Classification, and Regression) need to be combined with an algorithm. Many different algorithms can be used for DM. According to Kowalski (1979) is an algorithm based on two components. The first is a logical component used for solving the problem. The second one is the controlling component which determines the strategies one should choose to be able to solve the specific problem.

2.3.4.1 Artificial Neural Networks

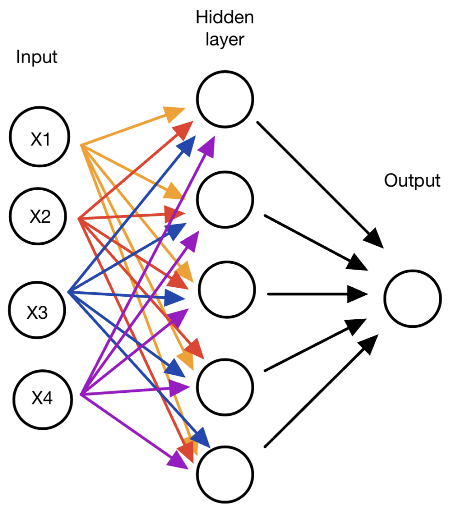
Artificial Neural Networks (ANN) is, according to Maind et al. (2014), a paradigm for processing information that takes inspiration from the biological nervous system. ANN is constructed by nodes/neurons which are connected together, making a complex network. Different layers organize these nodes. The complexity of how these nodes are placed, how many nodes, and how many layers the nodes can be placed in can be seen as the biggest challenges when setting up an ANN. In most cases of ANN, three layers are usually used: Input, hidden, and output, as seen in 2.10b. Several hidden layers can be constructed between the input and output layers. As said, artificial neurons are trying to imitate the biological neurons in the human body. Here the input layer will serve as the dendrites and the outputs as the synapse (Kukreja et al. (2016)). 2.10a visualize how a neuron is constructed. The input values are represented as x_i with accompanying weight W_i. The neuron receives a total input which is the sum of all the inputs multiplied with its weights (Krogh (2008)) (Abraham (2005)), see 2.7.

$$net = \sum_{N}^{i=1} w_i x_i = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n \tag{2.7}$$

The output F(net) 2.10a is called the activation function, of which there are many different types. These functions check if the net value exceeds a given threshold value θ. A common activation function is the Sigmoid function. This function will give an output of 1 if the net value exceeds the threshold value. Otherwise, the output will be 0. Other functions often vary from -1 to 1.



(a) Presentation of how a neuron is constructed Kukreja et al. (2016)



(b) Visual presentation of architecture of ANN

Figur 2.10: ANN

ANN is constructed to learn from past events like the human brain. This is done in two ways, either supervised or unsupervised. In supervised learning, the user will have both input and output values. The system is trained so that it manages to calculate the already given output value. This is done by running the process several times, where the output is adjusted against the error from the original output. The approximately correct output value is found by altering the different weights of the neurons. This process is called the back-propagation algorithm.

On the other side is unsupervised training. This is characterized by the fact that the user only knows the input values and not the output values. Thus the machine must find out how it will manage to organize the input data.

2.3.4.2 Decision Trees

Decision trees (DT) are another popular algorithm used for DM. It is categorized as a supervised machine learning algorithm and can be used for classification and regression. The model is built in a top-down recursive way (Zhong (2016)) and can resemble the structure of a flow chart. It starts from the top with a root node consisting of the entire data set. The data is further split up in the best way possible and is then placed in the associated node. The outcrops from the root node are called branches. The tree size is based on how many decision nodes or question nodes are used. These nodes classify the data. Ideally, you want to split the data so that it ends within its class. In many cases, separating the data into homogeneous classes will not be possible. In such cases, the separation that gives the lowest variance is chosen. The two methods that can be used to measure this variance are entropy and the Gini index (Kingsford and Salzberg (2008)). When the data has gone through enough decision nodes, it will finally reach the leaf node, representing a possible outcome.

Bagging and boosting are two methods that can be used to improve the prediction for classification or regression cases. These techniques can be adapted to decision trees and other models like ANN. Bagging is also called bootstrap aggregation and was developed by Breiman (1996). This method is based on dividing the data into bootstrap samples. Then a prediction method is used on the samples, for example, decision trees, and the overall result is combined (Sutton (2005)). As opposed to just looking at the result from one decision tree, the combined result reduces the variance, meaning a better prediction of the data. Random Forest (RF) is a DM algorithm that uses bagging. Here the data is split into several decision trees. These are chosen so that the correlation between the decision trees is as low as possible, resulting in a better prediction.

Another algorithm is Gradient Boosting (GB) which is used for regression (Bentéjac et al. (2021)). In this algorithm, it is common to use decision trees as prediction models. The approach is similar to bagging and aims to approximate the function of the data set, F(x). Boosting is an iterative method where a loss function, L(y, F(x)), is defined and decreased throughout the process. In other words, the algorithm will run until the error or residuals is as close to zero as possible. The process is started by making an initial prediction for the function. This can, for example, be the mean $\widehat{y}$ of the values. Based on the initial prediction, the decision trees will predict the next and improved function, where each decision tree corrects for the errors made by the previous tree. One problem with GB is over-fitting. This is because as the residuals approach their minimum, this will lead to a more complex prediction for every iteration.

2.3.4.3 Support Vector Machines

Support Vector Machine (SVM) is a classification algorithm. According to Noble (2006), SVM is divided into four concepts: The separating hyperplane, the maximum-margin hyperplane, the soft margin, and the kernel function. The separating hyperplane is a straight line that goes the high-dimensional space, with the purpose of separating the samples (Noble (2006)). How this hyperplane is drawn is the next step: The maximum-margin hyperplane. The line is placed so that it separates the different classes as much as possible in the middle. To be able to do this, some expression vectors are selected from each class. These are used to determine the final position of the line. Here, the largest distance between the points and the hyperplane is selected. On the other hand, it is not certain that the different classes can be separated perfectly by a linear hyperplane. In the real world, where data is more complex, the data points from the different classes will overlap. These values can often be seen as outliers, and it can therefore be relevant not to consider them when the hyperplane is to be drawn. This is a concept where the user can determine how many of these outliers will be allowed to cross the hyperplane to the opposite side of the class they belong to. The choice of the width of the soft margin requires the user to have sufficient knowledge. Both too-narrow and too-wide soft margins will lead to misclassifications. As said, splitting the data into separate classes is not always possible. This is to say that the data is nonseparable, even with the help of a soft margin. However, one method that can be used is the kernel function. These functions transform the data into another dimension. The data generally goes from a lower-dimensional space to a higher-dimensional space (Noble (2006)). There are many different kernel functions, meaning the user has to choose the correct one to get the best separation of the classes.

2.3.4.4 Multiple Regression

Multiple Regression (MR) is a technique that is widely used in statistics and many other fields. As seen in 2.8, the approach is composed of input variable or independent descriptive variables $(x_1, ..., x_n)$, the adjustable coefficients $(a_0, ..., a_n)$ and Y is the predicted value, or the dependent variable (Nakamura et al. (2017)) (Tinoco (2012)).

$$Y = a_0 + a_1 * x_1 + a_2 * x_2 + \cdots + x_n + a_n \tag{2.8}$$

2.3.5 Performance of the models

To be able to understand how well the performance of the specific model is, an evaluation measure can be used. Based on these, it is possible to determine how well the predicted equation fits the input values.

2.3.5.1 The Squared Correlation Coefficient, R^2

$$R^2 = \left(\frac{\sum(y_i - \overline{y}) * (\widehat{y_i} - \overline{\widehat{y}})}{\sqrt{\sum(y_i - \overline{y})^2 * \sum(y_i - \overline{\widehat{y}})^2}} \right)^2 \tag{2.9}$$

In 2.9, the equation for the squared correlation coefficient is given (Asuero et al. (2006)). R^2 is a useful evaluation measure and ranges from 0 to 1. The closer R^2 value is to 1, the better the assumed equation found from the regression analysis.

2.4 Comparison of results from articles in litterateur review

Author	Year	Name on article
Eyo, U	2022	Strength Predictive Modelling of Soils Treated with Calcium-Based Additives Blended with Eco-Friendly Pozzolans —A Machine Learning Approach
Tran, Van Quan	2022	Hybrid gradient boosting with meta-heuristic algorithms prediction of unconfined compressive strength of stabilized soil based on initial soil properties, mix design and effective compaction
Ngo, Huong Thi Thanh	2021	Application of Artificial Intelligence to Determined Unconfined Compressive Strength of Cement-Stabilized Soil in Vietnam
Yousefpour, Negin	2021	Stiffness and Strength of Stabilized Organic Soils—Part II/II: Parametric Analysis and Modeling with Machine Learning
Wang, O	2013	Preliminary Model Development for Predicting Strength and Stiffness of Cement-Stabilized Soils Using Artificial Neural Networks
Shrestha, Rakshya	2012	Development of Predictive Models for Cement Stabilized Soils
Das, Sarat Kumar	2011	Application of Artificial Intelligence to Maximum Dry Density and Unconfined Compressive Strength of Cement Stabilized Soil
Tinoco, Joaquim	2011	Application of data mining techniques in the estimation of the uniaxial compressive strength of jet grouting columns over time

Table 2.3: Summary of articles used to study data mining techniques applied to geotechnical engineering

Reference	Data mining technique				Preformance of the model R^2		
	ANN	SVM	DT	MR	UCS	E	MDD
Eyo et al. (2022)			X		0.900	-	-
Tran (2022)			X		0.966	-	-
Ngo et al. (2021)	X	X	X		0.925	-	-
Yousefpour et al. (2021)	X			X	>0.9	>0.8	-
Wang and Al-Tabbaa (2013)	X				0.932	0.804	-
Shrestha and Al-Tabbaa (2012)	X				-	-	-
Das et al. (2011)	X	X			0.880	-	0.910
Tinoco et al. (2011)	X	X		X	0.940	-	-

Table 2.4: Visualisation of the different Data Mining techniques used and the performance of the model. The R^2 value given applies to the model that gave the best result. Under performance of the model, the acronyms are given as:
UCS = Unconfined Compressive Strength
E = Young's modulus (Stiffness)
MDD = Maximum Dry Density

In 2.3, the articles that are used to look into data mining techniques are listed. All these articles are about ground stabilization, either deep mixing (wet or dry) or jet grouting. The main goal of the various articles is to find an alternative to laboratory tests to determine the mechanical properties of the stabilized soil. This is because laboratory tests often are expensive and timeconsuming. Different algorithms are used in the articles, which are shown in 2.4. From this table, it can be seen that the algorithm that is used the most times in the given articles is ANN. The DT and SVM models are also central in some of the articles, compared to MR, which is mostly used as a basic model. Based on the performance of the models, shown in 2.4, the algorithms give reliable results. There are some differences in performance, but that is to be expected due to the differences in data mining technique, Dataset, and input variables. The latter two are shown in 2.5 and 2.6. Here there are large differences in the size of the various data sets as well as differences in which input values have been chosen. In summary, most of the input variables are related to the characteristics and conditions of the soil. On the other hand, the execution parameters are not well represented among the input variables in 2.5 and 2.6.

Reference	Dataset	Input	Output
Wang and Al-Tabbaa (2013)	Two datasets: One consisted of 219 data cases of laboratory-prepared inorganic silty clays stabilized by either dry or wet mixing using Portland cement. The second one was composed from 223 data cases of cement-mixed sands.	Gravel Content Sand Content Silt Content Clay Content Water Content Liquid Limit Plastic Limit Plasticity Index Liquidity Index Cement Content W/C Ratio Curing Time Curing Stress	E50 UCS
Shrestha and Al-Tabbaa (2012)	220 data cases of cement-stabilized soils collected from a number of cement deep mixing projects	Soil Water Content Sand Content Silt Content Clay Content Organic Matter Content Binder Dosage Age Curing Temperature	UCS
Das et al. (2011)	The database consists of soils from 29 different sites from Canberra	Liquid limit Plasticity index Clay content Sand content Gravel content Moisture content Cement content	MDD UCS
Tinoco et al. (2011)	175 results derived from 35 JG laboratory formulations	W/C ratio Type of cement Strength class of cement Kilograms of cement by cubic meter of soil Age og the mixture Specific weight of the sample Water content % of sand % of silt % of clay	UCS

Table 2.6: Overview of Dataset, Input, and Output

Reference	Dataset	Input	Output
Eyo et al. (2022)	A dataset of 392 soils stabilised using cementitious additives'-enriched agro-based pozzolans in various proportions and combinations, compacted and cured for 7, 14 and 28 days	Values of agro-based pozzolans Cementitious additives Soil class Liquid limit Plasticity index Plastic limit Curing duration Strength class	UCS
Tran (2022)	The database is collected from 111 soils samples from 29 rammed earth building sites in Canberra, Australia.	Liquid limit Plastic limit Plasticity index Linear shrinkage Clay content Sand content Gravel content Lime content Cement content Asphalt content Optimum moisture content Maximum dry density	UCS
Ngo et al. (2021)	A total of 216 soil–cement samples were mixed in the laboratory and compressed to determine the UCS	Soil type Moisture content Wet density of soil The soil sampling depth Amount of cement Specimen diameter Specimen length Specimen area Specimen volume Mass of specimen Density of specimen Curing condition Curing period Type of cement	UCS
Yousefpour et al. (2021)	1030 unconfined compression test on three organic soils. Irish Moss Peat and two organic clays	Organic content Water content The ratio of binder for Portland Cement The ratio of binder for Blast Furnace Slag The ratio of binder for Pulverized Fuel Ash The ratio of binder for Lime The ratio of Binder for Magnesium Oxide The ratio of Binder for Gypsum The quantity of binder Grout to soil ratio Water to binder ratio Sample size (diameter to height ratio) Aging (time) Temperature Relative humidity Carbonation	E50 UCS

Table 2.5: Overview of Dataset, Input, and Output

3 Presentation of Dataset

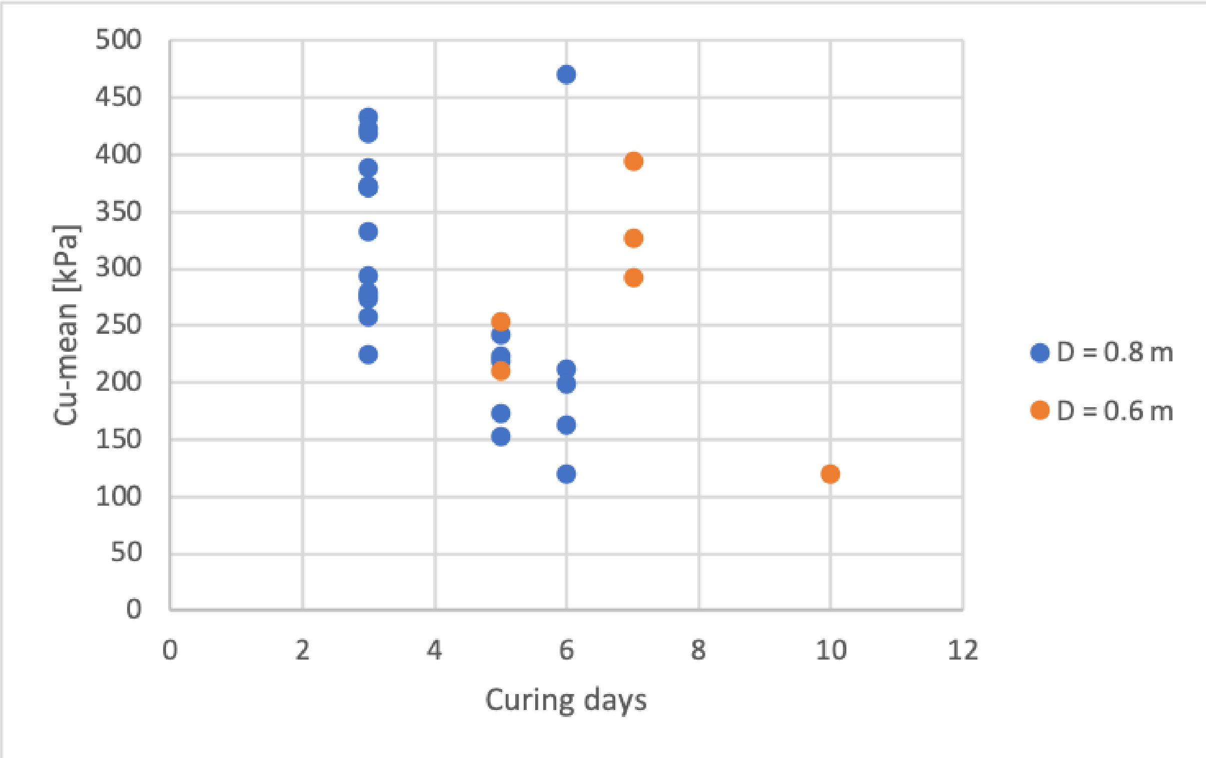
Column No.	Diameter [mm]	Binder dosage [kg/m^3]	BRN	CU-mean [kPa]	Age [Days]
1	800	58.3	250	372.81	3
2	800	61.5	245	257.39	3
3	800	58	251	371.33	3
4	800	60.4	242	277.02	3
5	800	53.1	250	273.84	3
6	800	57.5	248	432.62	3
7	800	58.6	249	331.75	3
8	800	53.2	244	293.34	3
9	600	61.5	237	119.16	10
10	600	56.9	235	394.72	7
11	600	62.8	235	292.55	7
12	600	57.2	237	326.77	7
13	600	59	236	210.22	5
14	800	61	250	223.40	5
15	800	60.6	250	241.82	5
16	800	61	250	153.34	5
17	800	61.7	245	211.72	6
18	800	61	251	198.39	6
19	800	59.7	250	119.23	6
20	800	59.6	245	219.46	5
21	800	60.9	250	173.52	5
22	800	64.9	226	418.85	3
23	800	61.5	241	388.31	3
24	800	61.1	222	423.48	3
25	800	61.1	222	224.26	3
26	800	60.2	238	163.20	6
27	600	58.4	233	253.97	5
28	800	62.7	221	278.93	3
29	800	64.3	262	370.68	3
30	800	60.8	248	469.78	6

Table 3.1: Summarization of Data provided from Keller Geoteknikk

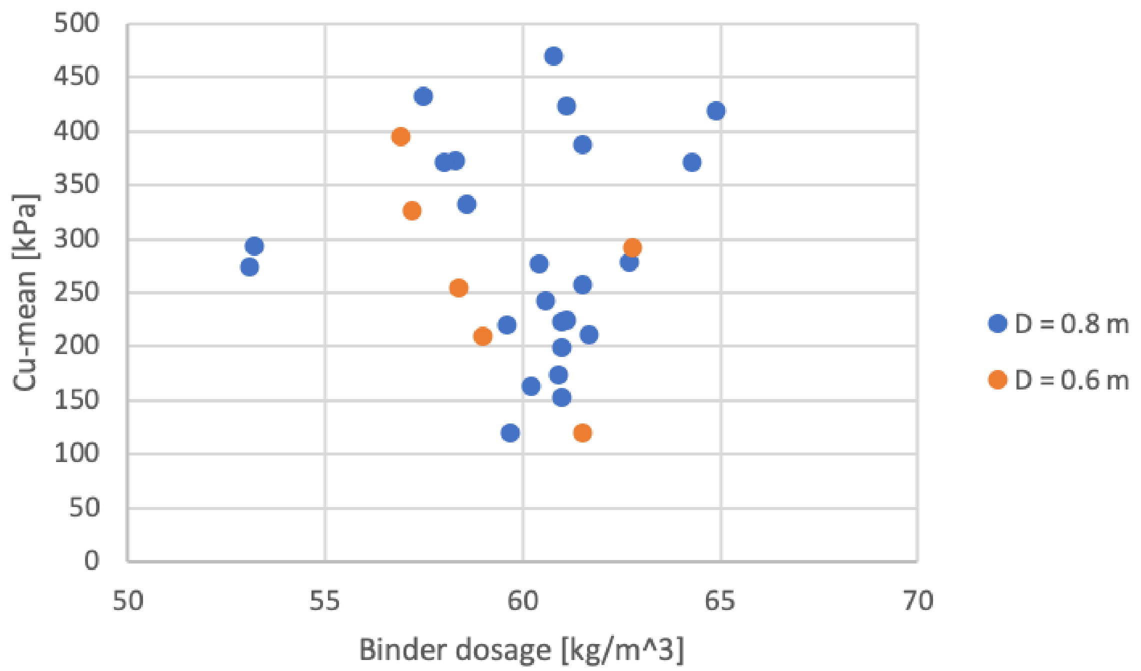
Keller Geoteknikk has provided data from 30 deep dry mixing columns, as seen in 3.1. The column numbers are named from 1-30, due to confidentiality. The diameter, Binder dosage and BRN were given in the chart logs, where the mean value is used. The Cu-mean and curing days were given from the FKSP results. The Cu-mean values seen in 3.1, were found by making numerical data from the graphs, provided from Keller, and finding the mean value of all these points.

4 Results

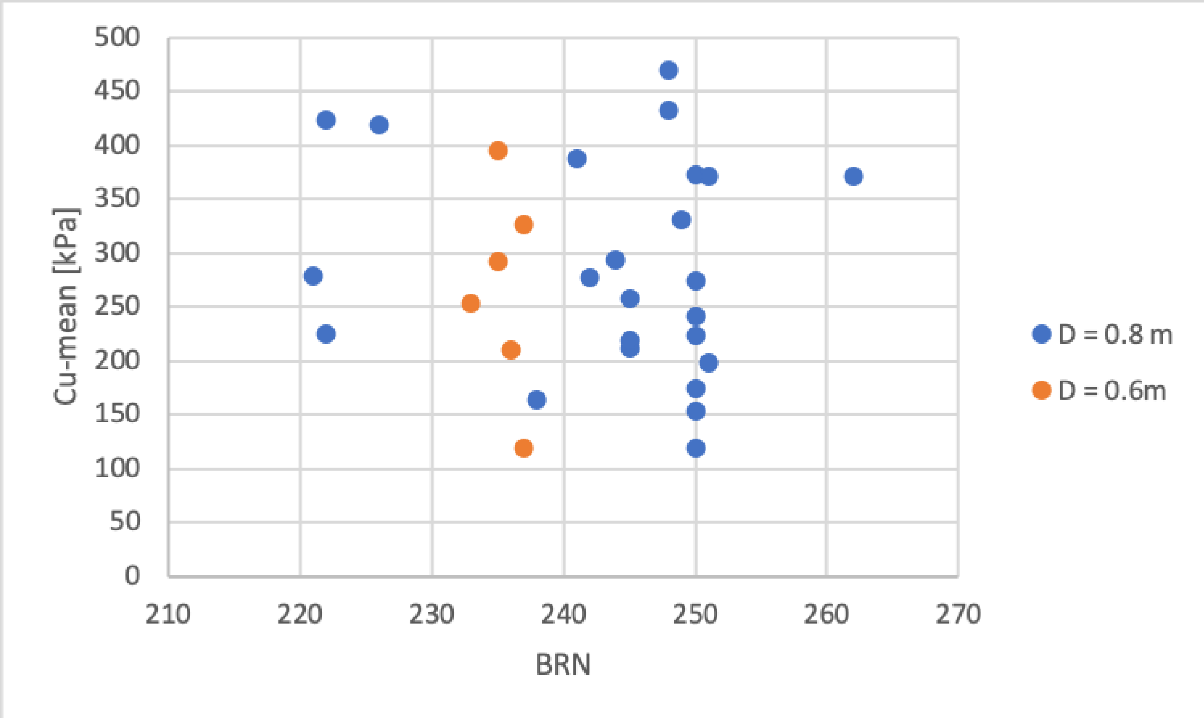
4.1 Scatter Plots and Correlation



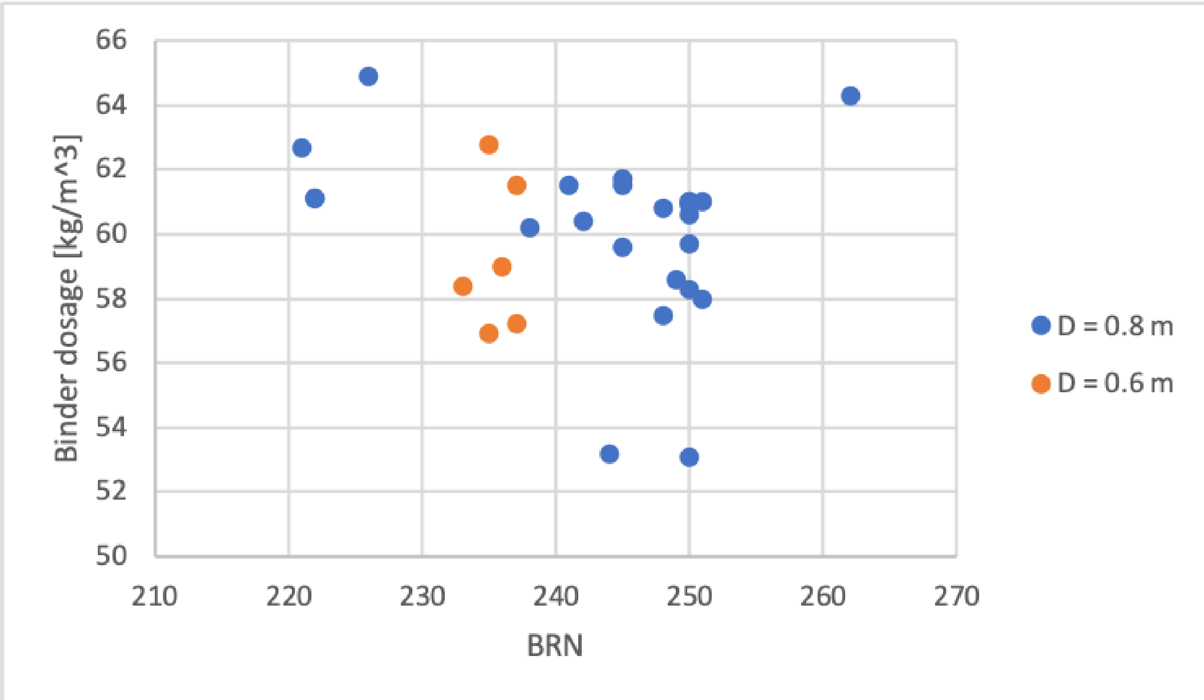
Figur 4.1: Curing days Vs Cu-mean



Figur 4.2: Binder dosage Vs Cu-mean



Figur 4.3: BRN Vs Cu-mean



Figur 4.4: BRN Vs Binder dosage

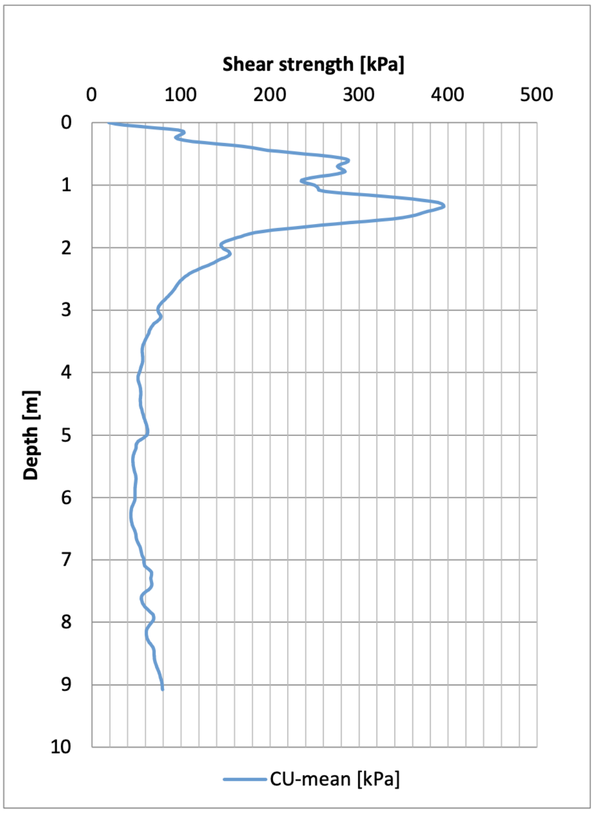
In this thesis, a preliminary analysis has been made. It has been chosen to base this analysis on BRN, Binder dosage, and Cu-mean. The executing parameters penetration rate, lifting speed, and RPM are determined not to include since these are included in the BRN. In 4.1, 4.2, 4.3 and 4.4, the scatter plots between Binder dosage, BRN, Cu-mean and curing days are shown.

None of the plots show any particular signs of connection between the data. This agrees with the correlation results in 4.1. Correlation is a statistical tool used to measure the relationship between two variables. There are three types of correlation, positive, negative, and zero. In this case, the correlation results vary from -0,0568 to -0,1851. The lowest correlation is between Binder dosage and Cu-mean. The result tends toward zero correlation, meaning no relationship exists between the variables. This also applies to the relationship between BRN and Cu-mean. However, there is a weak negative correlation between Binder dosage and BRN. This means that as one variable becomes larger, the other one will decrease.

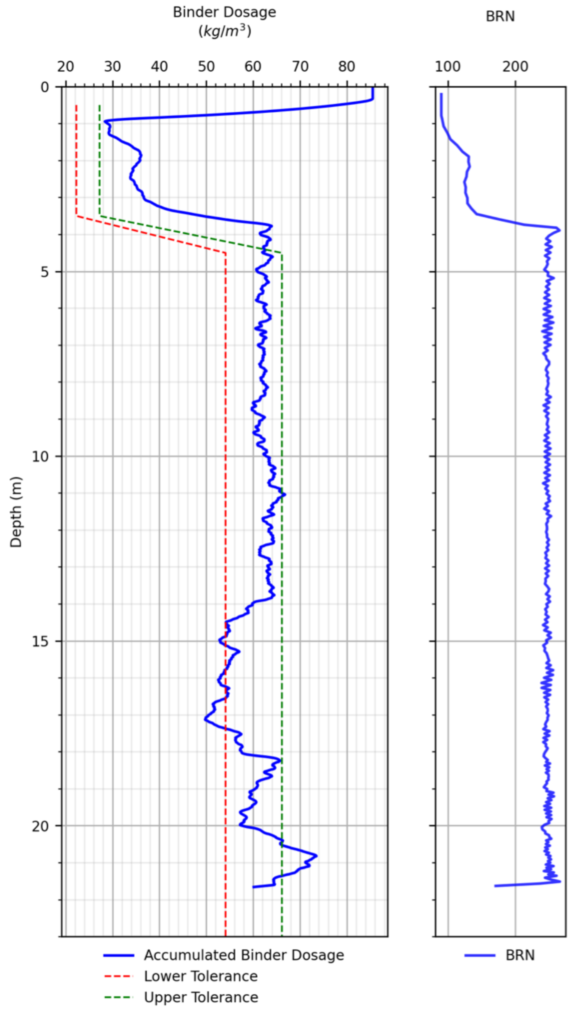
	CU-Mean	BRN	Binder dosage
CU- Mean	1	-0.0912	-0.0568
BRN	-0.0912	1	-0.1851
Binder dosage	-0.0568	-0.1851	1

Table 4.1: Correlation matrix between CU-mean, BRN, and Binder dosage

The weak correlation between the BRN and binder dosage against the Cu-mean can be seen in 4.5. In the upper layer, from 0 to -5 meter depth, a low binder dosage and BRN is used. Although one would think that this would result in lower shear strength, it can be seen in 4.5a that this is the shear strength is at its maximum at this depth. On the other hand, when the binder dosage and BRN are increased, the shear strength is reduced.



(a) Shear Strength



(b) Production log showing Binder dosage and BRN

Figur 4.5: Quality control for Column nr. 26

5 Summary and Conclusion

This project aims to see if it is possible to find patterns in the data provided from the DDM process related to the mechanical properties of the stabilized soil. The data are implemented into a DDM database before preliminary analyses are conducted. As seen in 4.1, there is a weak relationship between the variables. This means that it will not be possible to say precisely how the tested executing parameters are related to the shear strength of the finished column. This is because it has not been possible to detect reliable correlations between the variables. Based on what is found in the literature by Larsson, see 2.2, it was assumed that the BRN and binder dosage would greatly impact the strength of the stabilized product. According to Larsson (2005), both parameters significantly influenced the strength variability. As seen in 4.5, it is difficult to say anything precisely about the shear strength based on only the binder dosage and BRN. Here, the highest shear strength depths had the lowest values for BRN and binder dosage.

When going through the production logs and the results from FOPS testing, one notices that there are generally small variations in BRN and binder dosage, as seen in 2.2. On the other hand, there are large variations in shear strength, ranging from 119.16 kPa to 469.78 kPa. As seen in 4.5,

The machine monitor and records many different parameters but will not provide any data on the soil conditions. To obtain such data, additional soil investigation tests are needed. Since the data in the dataset are taken from different projects with different soil conditions, this entails great uncertainty. Compared to the BRN and binder dosage, the rheology of the soil, according to Larsson (2005), is an even more critical factor for the strength increase. This can be seen again in the 2.5 and 2.6, where it can be seen that input variables related to the rheology are essential for the predicted output value.

This is seen further in 4.5a, where it is natural to think that there is a top crust with a high shear strength with soft clay underneath due to the sudden drop in shear strength. By just looking at the graphs in 4.5b, it would be more natural to assume that one should have had a lower shear strength at the top of the column and a higher one at the bottom due to the increase in BRN and binder dosage. Here one sees the importance of having information about the soil conditions to predict the strength increase.

The problem is that the strength increase in cement-stabilized soil is a complicated process with many influencing factors. Therefore, it won't be easy to say something about the final strength of the stabilized column by only looking at the executing parameters. These are presented in 2.1, which are summarized into four main categories: characteristics of stabilizing agent, Characteristics and conditions of soil, mixing conditions, and curing conditions.

In this thesis, only a simple correlation analysis has been carried out. This is not as accurate as the complex Data Mining analyses carried out on the studies looked at in the litterateur study. Thus, it is possible that by using more advanced methods, it could have been achievable to find trends or patterns in the data.

5.1 Recommendation for Further Work

The work done in this thesis is the starting phase of what will be done in the following master's thesis next year. To provide reliable results, it is necessary to get data about the soil conditions. This can be seen from the results found in this thesis. When the data is obtained, it is added to the already-created database. This also needs to be updated with all the execution parameters so that it is not only Binder dosage, BRN, and Cu mean that are taken into account. More input variables can give a better result, but interpreting the data may be more challenging. Therefore, data mining algorithms that have been considered in this thesis will be used.

Bibliography

Abraham, A. (2005). Artificial neural networks. *Handbook of measuring system design*.

Asuero, A. G., Sayago, A., and González, A. (2006). The correlation coefficient: An overview. *Critical reviews in analytical chemistry*, 36(1):41–59.

Azevedo, A. and Santos, M. F. (2008). Kdd, semma and crisp-dm: a parallel overview. *IADS-DM*.

Bentéjac, C., Csörgő, A., and Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54(3):1937–1967.

Boutouil, M. and Levacher, D. (2005). Effect of high initial water content on cement-based sludge solidification. *Proceedings of the Institution of Civil Engineers-Ground Improvement*, 9(4):169–174.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.

Das, S. K., Samui, P., and Sabat, A. K. (2011). Application of artificial intelligence to maximum dry density and unconfined compressive strength of cement stabilized soil. *Geotechnical and Geological Engineering*, 29(3):329–342.

Eyo, E. U., Abbey, S. J., and Booth, C. A. (2022). Strength predictive modelling of soils treated with calcium-based additives blended with eco-friendly pozzolans—a machine learning approach. *Materials*, 15(13):4575.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996a). From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37–37.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996b). The kdd process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11):27–34.

Ghee, C. K. (2006). Constitutive behaviour of cement treated marine clay.

Huawen, X. (2009). Yielding and failure of cement treated soil.

Keller (2022). Dry soil mixing. (Accessed on 04/11/2022).

Kingsford, C. and Salzberg, S. L. (2008). What are decision trees? *Nature biotechnology*, 26(9):1011–1013.

Kowalski, R. (1979). Algorithm= logic+ control. *Communications of the ACM*, 22(7):424–436.

Krogh, A. (2008). What are artificial neural networks? *Nature biotechnology*, 26(2):195–197.

Kukreja, H., Bharath, N., Siddesh, C., and Kuldeep, S. (2016). An introduction to artificial neural network. *Int J Adv Res Innov Ideas Educ*, 1:27–30.

Larsson, S. (2005). State of practice report–execution, monitoring and quality control. *Deep Mixing*, 5:732–785.

Larsson, S. (2014). Mixing processes for ground improvement by deep mixing. *Svensk Djupstabilisering*.

MacLaren, D. C. and White, M. A. (2003). Cement: Its chemistry and properties. *Journal of Chemical Education*, 80(6):623.

Maind, S. B., Wankar, P., et al. (2014). Research paper on basic of artificial neural network. *International Journal on Recent and Innovation Trends in Computing and Communication*, 2(1):96–100.

Mariscal, G., Marban, O., and Fernandez, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review*, 25(2):137–166.

Nakamura, K., Yasutaka, T., Kuwatani, T., and Komai, T. (2017). Development of a predictive model for lead, cadmium and fluorine soil–water partition coefficients using sparse multiple linear regression analysis. *Chemosphere*, 186:501–509.

Ngo, H. T. T., Pham, T. A., Vu, H. L. T., and Giap, L. V. (2021). Application of artificial intelligence to determined unconfined compressive strength of cement-stabilized soil in vietnam. *Applied Sciences*, 11(4):1949.

Noble, W. S. (2006). What is a support vector machine? *Nature biotechnology*, 24(12):1565–1567.

Olson, D. L. and Delen, D. (2008). *Advanced data mining techniques*. Springer Science & Business Media.

Piatetsky, G. (2014). Crisp-dm, still the top methodology for analytics, data mining, or data science projects. (Accessed on 27/10/2022).

Pradeep, G. and Vinu, T. (2015). Effect of organic matter on the geotechnical properties of soil and impact of lime-salt stabilization in strength improvement of organic soil. *International Journal of Engineering Research and Technology*, 3(29).

Rocscience (2022). Spatial variability. (Accessed on 06/12/2022).

Santos Barros, A. A. (2019). Dry deep soil mixing soil-cement column panels as bottom struts for excavation support: Revising of design methodology in scandinavia.

Shrestha, R. and Al-Tabbaa, A. (2012). Development of predictive models for cement stabilized soils. In *Grouting and Deep Mixing 2012*, pages 221–230.

Sutton, C. D. (2005). Classification and regression trees, bagging, and boosting. *Handbook of statistics*, 24:303–329.

Terashi, M. (1997). Theme lecture: Deep mixing method-brief state of the art. In *Proc. 14th ICSMFE*, volume 4, pages 2475–2478.

Timoney, M. J. and McCabe, B. A. (2017). Strength verification of stabilized soil–cement columns: a laboratory investigation of the push-in resistance test (pirt). *Canadian Geotechnical Journal*, 54(6):789–805.

Tinoco, J., Correia, A. G., and Cortez, P. (2011). Application of data mining techniques in the estimation of the uniaxial compressive strength of jet grouting columns over time. *Construction and Building Materials*, 25(3):1257–1262.

Tinoco, J. A. B. (2012). Application of data mining techniques to jet grouting columns design.

Tran, V. Q. (2022). Hybrid gradient boosting with meta-heuristic algorithms prediction of unconfined compressive strength of stabilized soil based on initial soil properties, mix design and effective compaction. *Journal of Cleaner Production*, 355:131683.

Wang, O. and Al-Tabbaa, A. (2013). Preliminary model development for predicting strength and stiffness of cement-stabilized soils using artificial neural networks. In *Computing in Civil Engineering (2013)*, pages 299–306.

Yousefpour, N., Medina-Cetina, Z., Hernandez-Martinez, F. G., and Al-Tabbaa, A. (2021). Stiffness and strength of stabilized organic soils—part ii/ii: Parametric analysis and modeling with machine learning. *Geosciences*, 11(5):218.

YUTAO, P. (2016). Effective stress random finite element analysis of cement-treated clay ground.

Zhang, C., Yang, J., Ou, X., Fu, J., Xie, Y., and Liang, X. (2018). Clay dosage and water/cement ratio of clay-cement grout for optimal engineering performance. *Applied Clay Science*, 163:312–318.

Zhong, Y. (2016). The analysis of cases based on decision tree. In *2016 7th IEEE international conference on software engineering and service science (ICSESS)*, pages 142–147. IEEE.