Proceedings of the XVII ECSMGE-2019
Geotechnical Engineering foundation of the future
ISBN 978-9935-9436-1-3

doi: 10.32075/17ECSMGE-2019-0617

XVII European Conference
ECSMGE 2019
Reykjavik
on Soil Mechanics
and Geotechnical Engineering

Machine learning to detect sensitive materials with CPTu in Norway

Apprentissage automatique de la détection de matériaux sensibles avec CPTu en Norvège

S.M. Valsson
Norwegian Public Roads Administration

ABSTRACT: Sensitive materials are of great concern in geotechnical design in Norway, as failures involving such materials can potentially have very serious consequences. Knowing where these materials are found is therefore essential for safe and economical geotechnical design. The CPTu test has over the last two decades become a popular soil investigation method in Norway. Since early 2017, the Norwegian Public Roads Administration has been systematically registering which road-projects contain CPTu- and laboratory test results in the same boreholes. This work has produced a set of labeled data from over 240 CPTus.

The only generally accepted method of detecting highly sensitive materials is laboratory testing on soil samples, but various methods for interpreting in situ tests have been developed over the years to indicate if these kinds of materials are present. As soil sampling and laboratory testing are both expensive and time consuming, in-situ methods that reduce the need for or better focus the sampling are valuable.

This study describes how a machine learning algorithm can be trained with previous CPTu and laboratory results in order to detect highly sensitive materials in situ.

RÉSUMÉ: Les matériaux sensibles sont un sujet de préoccupation majeur dans la conception géotechnique en Norvège, car les défaillances impliquant de tels matériaux peuvent avoir des conséquences très graves. Savoir où se trouvent ces matériaux est donc essentiel pour une conception géotechnique sûre et économique. Le test CPTu est devenu au cours des deux dernières décennies une méthode d'étude du sol en Norvège. Depuis début 2017, l'administration publique norvégienne des routes publiques enregistre systématiquement les projets de route contenant les résultats des tests CPT et de laboratoire dans les mêmes trous de forage. Ce travail a produit un ensemble de données étiquetées provenant de plus de 240 CPTus.

Les tests de laboratoire sur des échantillons de sol sont la seule méthode généralement reconnue pour détecter les matériaux très sensibles. Toutefois, diverses méthodes d'interprétation des tests in situ ont été développées au fil des années pour indiquer si ce type de matériaux est présent. L'échantillonnage de sol et les analyses de laboratoire étant à la fois coûteux et longs, les méthodes in situ qui réduisent la nécessité ou la focalisation de l'échantillonnage sont utiles.

Cette étude décrit comment un algorithme d'apprentissage automatique peut être formé avec des résultats antérieurs de CPTu et de laboratoire afin de détecter in situ des matériaux très sensibles.

Keywords: Soil investigations; Sensitive materials; CPTu; Machine learning

1 INTRODUCTION

The frequency of landslides involving sensitive materials in the last decade in Norway seems to be increasing when compared to historical records, where at least 8 landslides have been recorded during the last 10 years.

Some of these landslides have been initiated by small failures, thought to be caused by ground works, which then developed into larger landslides.

Even though this increase in may be inaccurate (earlier records may be lacking) the general consensus is that the current rate is unacceptable.

A first step in reducing the number of landslides in sensitive materials is to document where these materials are found and placing restrictions on developments in those areas in zoning plans. The goal of this study is to create a model that can help identify positions where sensitive materials are found.

2 SOIL INVESTIGATIONS

A common practice for soil investigations has been established in Norway over the last few decades. This is to a large extent due to the work done by the Norwegian geotechnical society, where experiences from different soil investigation techniques have been shared and new methods have been developed to meet the requirements of that period.

Over the last two decades the cone penetration test with porepressure measurements (*CPTu*) has become one of the most popular soil investigation methods. This is due to the fact that the test is highly repeatable (Lunne et al. 1997), quick to execute and provides large amounts of relevant data to interpret soil parameters compared to other field methods.

In order to transform experience from one site to the next the Norwegian Public Roads Administration (*NPRA*) has gathered CPTu- and corresponding laboratory data to a large table for further studies. We refer to this table as the *training data* in this paper.

2.1 Depth correction for CPTu

At any given time during a CPTu test, the main sensors are at different depths. This is due to geometry of the cone.

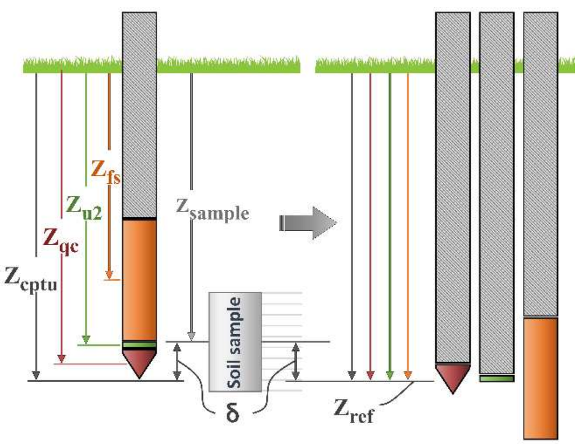

Figure 1. Defining the depth of CPTu registrations

This causes each of the registrations to be at a slight offset from the desired depth. In order to correct this, each of the registrations is shifted to a common reference depth, z_{ref}, as shown in *Figure 1*.

2.2 Sets of data

The goal of the study is to create a model to distinguish between sensitive and non-sensitive data, therefore points where relevant laboratory tests have been extracted from the training data.

This study uses classifiers, so points belonging to relevant groups are filtered from the rest of the training data with the criteria shown in Table 1.

Table 1. Point group definitions for this study

Point group	Definition
Quick clays	$c_{ur} < 0.5\ kPa$
Highly sensitive	$c_{ur} \leq 2.0\ kPa$ $and\ S_t > 15$
Not sensitive	$c_{ur} > 2.0\ kPa$

The remoulded undrained shear strength, c_{ur} (kPa), and sensitivity, S_t (-), in *Table 1* have been estimated using the swedish fallcone test on undisturbed and remoulded soil samples.

The three groups are of uneven sizes, where the two groups with sensitive points contain about 1/3 of the points when combined, and the not sensitive group countains about 2/3 of the points.

Laboratory tests are performed on 10cm pieces of undisturbed samples with a diameter of 54mm. Due to the logging interval of the CPTu cones, each 10cm soil sample in the training data has 5-10 corresponding CPTu registrations.

To remove redundant data and reduce scatter in the training data, the average of the CPTu registrations for each 10cm soil sample is calculated.

3 MACHINE LEARNING

Machine learning (*ML*) is a field within computer science which describes creating programs that define models (*learn*) directly from data. These models can be used to make predictions on new data.

Although this field has existed for decades, it is currently expanding rapidly. This is due to more data becoming available for analysis (systematic archiving), and modern computers having greater computing power.

It is important to remember that ML models rely solely on data, and using advanced algorithms does not compensate for data of low quality.

This study uses one of the simplest ML algorithms, k Nearest Neighbors classifier (*kNN*), to try to identify sensitive materials from non-sensitive materials using the CPTu test. This work relies on a machine learning framework for Python called Scikit-learn (Pedregosa et al. 2011).

3.1 k Nearest Neighbors classifier

The kNN algorithm is a geometric classifier which assigns a class to a point based on distances.

In order to use this algorithm a set of known points is required (*training data*), with both point values of some different variables (*features*) and identifying types (*labels*).

To classify a test point, distances to all points in the training data is calculated, and the list is sorted in ascending order.

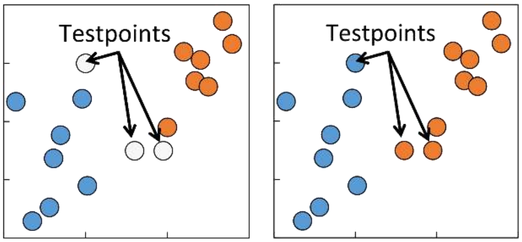

Figure 2. kNN classification of a fictional data (k=1)

The label of the majority of the *k* nearest points is assigned to the test point. Calculating the distance between points transforms the classification problem from *n* dimensions to just one, and the only tuning parameter is how many neighbors (*k*) to take into account.

Figure 2 shows how labels are assigned to fictional points based on the label of a single nearest neighbor (*k*=1).

kNN is a simple method, and easy to implement, but the features must be of comparable sizes for it to work. A unit increase in any of the features must have the same effect on the distance metric. It is therefore important to normalize (*scale*) the data before classification (Kotsiantis et al. 2006).

3.2 Scaling data

The test points will eventually be compared directly to the training points to calculate distances and it is therefore important that exactly the same transformation is applied to both sets.

$$x' = \frac{x-\bar{x}}{\sigma_x} \tag{1}$$

Where x (*any unit*) is the vector of original values, $\bar{x}$ is the average value of x and σ_x is the standard deviation of x. The resulting vector x' (*unitless*) has the average value 0 and standard deviation 1.

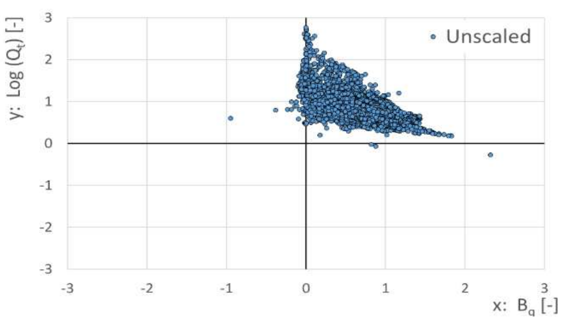

Figure 3. unscaled points from the training data

The average value and standard deviation refer only to the training data, the same exact values are then used to transform the test data.

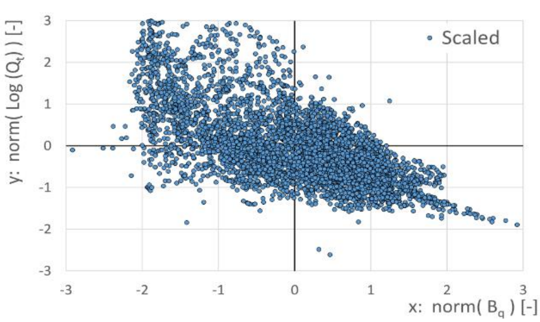

Figure 4. Scaled points from the training data

Figure 3 and Figure 4 are created using the same points, but eq (1) has been used to scale the points in Figure 4.

3.3 Cross validation and custom metric

It is important to evaluate if a particular model generates good results, and how it will generalize to new data.

Each model is therefore tested with known points, but it is important that the test points are not present in the training data. This is achieved by splitting the training data into a test- and (*new*) training set.

This study uses the Leave One Group out (*LOGOS*) train-/test split method, where points belonging to one specific group are used to define the test set and the rest of the groups are used to define model. In this case the groups are defined as the projects in the training data. This is shown in Figure 5.

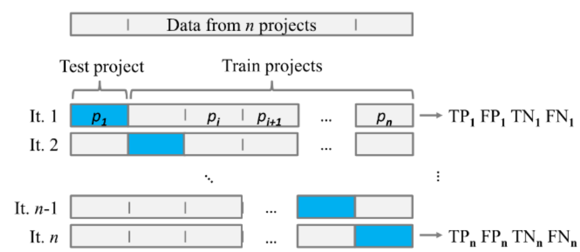

Figure 5. Shows the LOGOS cross validation split

The true positive rate is defined as

$$TPR = \frac{\sum TP}{\sum TP + \sum FN} \tag{2}$$

Where $\sum TP$ (-) and $\sum FN$ (-) are the sum of sensitive points from all LOGOS iterations that are classified as sensitive and not sensitive respectively.

The false positive rate is defined as

$$FPR = \frac{\sum FP}{\sum FP + \sum TN} \tag{3}$$

Where $\sum FP$ (-) and $\sum TN$ (-) are the sum of not sensitive points from all LOGOS iterations that are classified as sensitive and not sensitive respectively.

A unitless custom metric is defined as

$$CM = TPR - FPR \quad (4)$$

This is done in order to apply even weights to the sensitive and not sensitive points, while tuning the kNN model.

3.4 Feature selection

Feature selection describes the process of selecting relevant features to solve the problem at hand. Our training data contains 26 features, where many are similar (redundant), as well as other features not relevant for this task (depth/in-situ vertical pressure/…).

Recursive feature selection (*RFS*), is the process of iterating over all features, preforming cross validation (here LOGOS) during each iteration, and selecting the one that gives the best score or score increase (here CM). This process is repeated until the desired number of features has been selected

With only 26 features, a brute force feature selection (*BFFS*) is feasible. BFFS tests all possible combinations of 2 features and selects the pair with the best score.

Using all available features, the number of neighbors with the highest custom metric is identified as *k*=23. This is used to identify the best combination of features with *BFFS*. Repeating the search for *k* with this feature pair reveals a k_{opt}=47. This model is used to select 4 more features with *RFS*, the results are shown in Table 2.

Table 2. Results from feature selection using k=47

Feature	Iteration	TPR	FPR	CM
Bq (-)	Best pair	0,48	0,25	0,23
Rf (%)		0,56	0,07	0,49
q_e (kPa)	1	0,57	0,08	0,50
Fr (%)	2	0,57	0,07	0,50
f_{sn} (-)	3	0,56	0,07	0,49
f_s (kPa)	4	0,54	0,06	0,48

Table 2 shows that the maximum values for CM are reached when the model is defined with 3 variables, and using more features does not add value to the interpretation (but rather confuses it). It should be noted that adding the third variable only marginally improves the score.

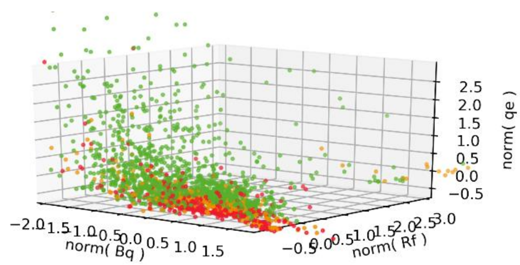

Figure 6. Points used to define the 3D kNN model from Table 2, green points are not sensitive, yellow points are highly sensitive and red points are quick clay.

Figure 6 shows the normalized points used to define the 3D kNN model that gave the highest CM for separating sensitive points from not sensitive points.

3.5 Visualizing the model

Visualizing models can give a better understanding of how they are likely to perform. Only the best pair of features are used to create images of decision boundaies (2D).

The number of neighbors is varied once more to see if adding the third feature had an impact on the optimal number of neighbor in the model, the results are shown in *Figure 7*.

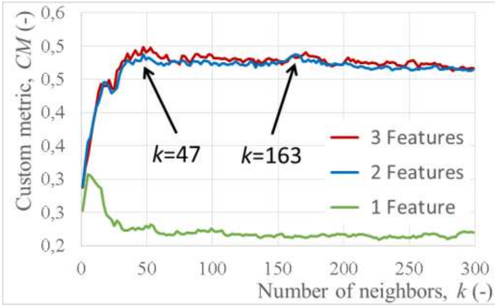

Figure 7. Custom metric for 1, 2 and 3 feature models as a function of the number of neighbors, k.

Figure 7 shows that optimal values for k for 2 features can be selected as either 47 or 163, and that 47 is also the optimal value when using 3 features. The point clouds and corresponding decision boundaries for kNN models using k=47 and k=163 are shown *Figure 8* and *Figure 9*.

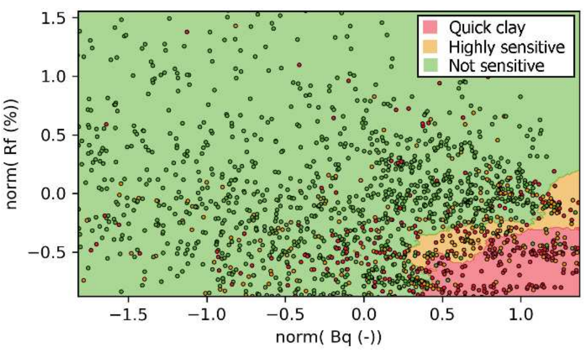

Figure 8. 2D projection of the points from Figure 6 and the decision boundary for a kNN model using 47 neighbors.

The decision boundaries shown in *Figure 8* and *Figure 9* show that increasing the number of neighbors to consider makes the decision boundaries less flexible (*more biased*), as each point in the training set has less influence on the result.

This can be either positive or negative, depending on the quality of the training data and the intended function of the model.

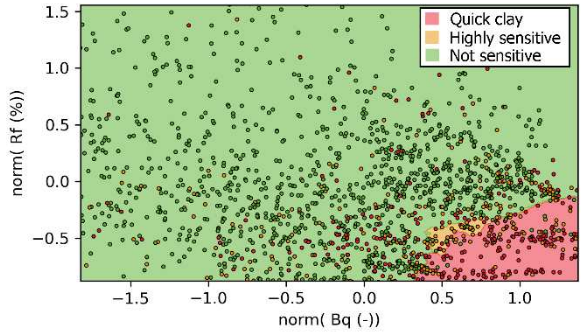

Figure 9. 2D projection of the points from Figure 6 and the decision boundary for a kNN model using 163 neighbors.

It is clear from *Figure 6*, *Figure 8* and *Figure 9* that there is large scatter in the training data, and many points from sensitive materials are being ignored in the depicted models. This tells us that one should not expect a perfect classification when applying the model to new data.

4 TESTING MODEL ON PROJECTS IN NORWAY

To test the model, three test positions have been selected from three different sites in Norway. These positions are not present in the training data and in addition to CPTu tests, the soil conditions are documented with rotary pressure soundings and laboratory tests on undisturbed soil samples.

The model produces a color coded vertical profile (shown on the right in *Figure 10*-*Figure 12*) where green indicates not sensitive, cyan indicates quick clay, and magenta indicates highly sensitive.

4.1 Test position 1 - Skatval

Skatval is located in Trøndelag (central norway), and the presence of sensitive materials has been documented in this area.

The results in *Figure 10* show that the kNN model identifies sensitive materials, but not at the same depths that the laboratory tests have shown.

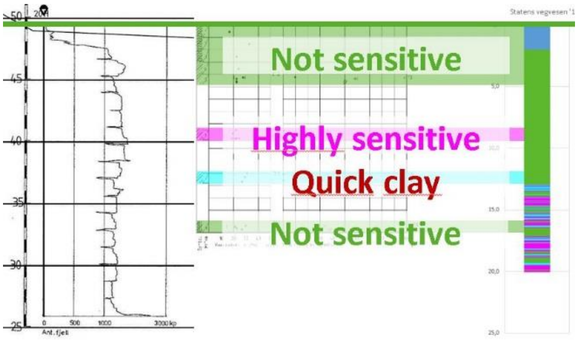

Figure 10. Rotary pressure sounding, laboratory tests and CPTu classification with kNN at Skatval.

Even though the model does not register sensitive materials at 8-9 m and 12-13 m depths, it does show a trend to not sensitive materials at 17-18 m depth.

The confidence of the classification *(% of the k-neighbours indicating a single class*) varies in this test position, and below the depth of 5 meters, no one class has more than 80% of the neighbors.

4.2 Test position 2 - Åby

Åby is located in Telemark (South Norway), and the soil investigations from this area were conducted in connection to a quick clay slide.

The soil conditions can be described as homogenuous sensitive clays below the depth of 4,5 meters.

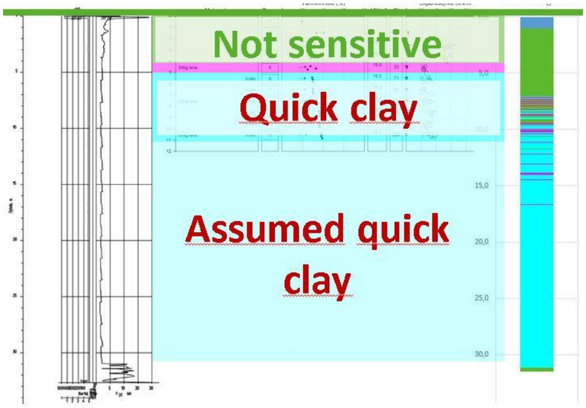

Figure 11. Rotary pressure sounding, laboratory tests and CPTu classification with kNN at Åby.

The proposed model indicates the presence of sensitive materials below the depth of 7,5 – 8 meters, and does therefore not pick up on the first 3-4 meters of sensitive materials.

In contrast to the Skatval test, this classification predicts with high confidence and between 17,5-30 meters the confidence is 100%.

4.3 Test position 3 - Rissa

Rissa is well known in the geotechnical community because of the quick clay landslide that occurred (and was filmed) in 1978.

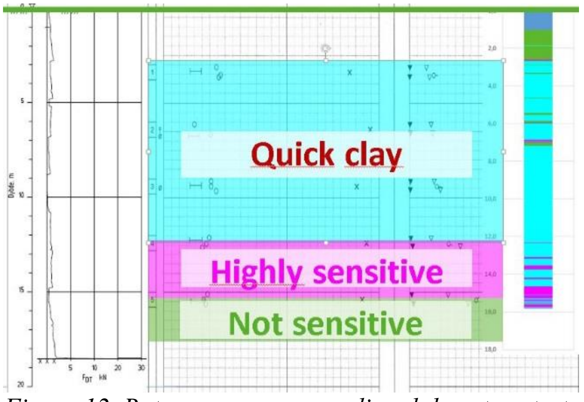

Figure 12. Rotary pressure sounding, laboratory tests and CPTu classification with kNN at Rissa.

The results from this test show that the proposed model indicates sensitive materials below the depth of 2,5 meters. Laboratory tests have not been carried out on samples above 2,5 meters and it is therefore unclear if the model correctly predicted the boundary between sensitive and not sensitive materials in this position.

It is worth noting that the model does indicate a transition from quick clay to highly sensitive materials at 15,3 meters depth, which is in accord with the laboratory results.

The classification confidence is about 80-100% for the sensitive materials, and about 60% for the not sensitive materials

5 CONCLUSIONS

This work has shown that a tuned ML classifier that is trained with selected features from projects in Norway is able to detect the presence of sensitive materials with the CPTu.

The precision of the classification is not perfect, but this is to be expected when compared the scatter of the training points in *Figure 8* and *Figure 9*.

One of the goals of this study was to implement the findings into an Excel spreadsheet so that the geotechnical engineers in the NPRA could implement these methods in future projects. This was possible, and the tool was used to generate the classification columns in *Figure 9* to *Figure 12*.

ML algorithms are exremely powerful, and can sometimes be used to create models that generate good predictions, but all of them rely on data. Using data of poor quality (or irrelevant to the problem at hand) will not generate good results.

The author believes that better models can be created for identifying sensitive materials with the CPTu (even using these exact same methods). The fastest path towards that goal is to verify the quality of the data in the training set, and weeding out soundings of low quality. Adding more data of high quality should also be a priority. The quality of the sounding that is being tested is of course equally important.

Taking the the scatter in the training data into account, it is not likely that using other features/ or implementing more advanced algorithms will yield vastly greater results for solving this problem with this training set.

Detecting sensitive materials with the CPTu is only one of many topics that can be analized with these kinds of methods. Soil type classification could be solved in the exact same manner. Interprating mechanical parameters such as undrained shear strength and friction angle would also be a valuable extension to this study.

6 ACKNOWLEDGEMENTS

The author would like to thank the many geotechnical engineers working in the NPRA. Without their efforts in gathering CPTu and laboratory data from actual projects, this work could not have been carried out.

The author would specially like to acknowledge Eigil Haugen, Samson Degago and Christian Berg Skjetne for their collaboration and discussions about this topic.

7 REFERENCES

Lunne T., Robertson P.K., Powell J.J.M. 1997. Cone penetration testing in geotechnical practice. London: Spon Press.

Kotsiantis S.B., Kanellopoulos D., Pintelas P.E. 2006. Data preprocessing for supervised learning. International Journal of Computational science, Vol. 1.

Pedregosa F., Varoquaux G., Gramfort A., Michael V., Thirion B., Grisel O., Blondel M. Prettenhofer P., Weiss R., Dubourg V. Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay É. 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, volume12.