



Austin, Texas, USA | August 16-19, 2020

ISFOG 2020 Data science prediction event

1. INTRODUCTION

Data science techniques are rapidly transforming businesses in a broad range of sectors. While marketing and social applications have received most attention to date, geotechnical engineering can also benefit from data science tools that are now readily available.

In the context of the ISFOG2020 conference in Austin, TX, a prediction event is launched which invites geotechnical engineers to share knowledge and gain hands-on experience with machine learning models.

This document outlines the scope of the ISFOG2020 data science prediction event which formulates pile driving predictions as a machine learning problem.

2. PROBLEM STATEMENT

2.1 Prediction problem

The problem proposed for the prediction event is the prediction of blowcount vs depth for jacket piles in North Sea soil conditions. In current engineering practice, blowcounts are derived using a three-step approach:

- Predict the soil resistance vs depth from geotechnical data (Alm and Hamre, 2001; Stevens et al., 1982);
- Calculate a bearing graph from 1D wave equation analysis for selected depths (Smith, 1960);
- Interpolate the bearing graphs to derive blowcounts.

This is a fairly complicated approach to relate soil properties, pile dimensions and hammer settings to the target variable, being the blowcount.

This prediction event will explore whether machine learning models can be trained on a set of pile driving records to provide reliable estimates of blowcount for unseen locations.

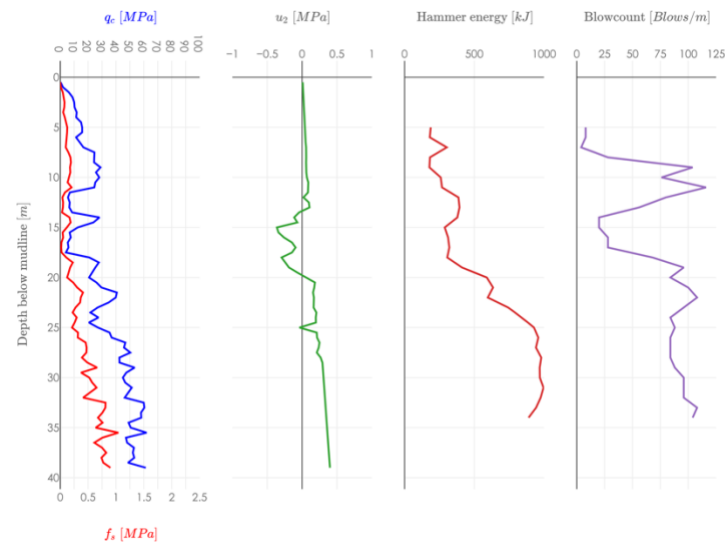
2.2 Data set

The available dataset consists of installation records for 120 piles in the North Sea and contains the following data sources:

- CPT profiles with cone tip resistance, sleeve friction and pore pressure measurements;
- Pile outer diameter and wall thickness;
- Blowcount and hammer energy.

To facilitate participation and keep a focus on the data science techniques, the data has already been processed (Figure 1) and mapped to a regular grid with nodes every 0.5m.

Figure 1: Results of data processing and grid mapping routine for one location



For the prediction event, the data set is partitioned into a training dataset of 100 piles and a validation data set of 20 piles. Only the training data will be provided, the validation set will be used to assess the quality of participant's models.

Participants can download the dataset in .csv format from the GitHub project page:

[\(https://github.com/isfog/datascience2020/\)](https://github.com/isfog/datascience2020/).



Austin, Texas, USA | August 16-19, 2020

3. GUIDELINES FOR PARTICIPATION

3.1 Target audience

This prediction event targets individuals or groups with an interest in data science techniques. Although some knowledge of the techniques can be beneficial, we also invite people who have no prior experience with machine learning.

To facilitate participation, a tutorial which demonstrates the basics of getting started with data science and building a machine learning model for the pile driving problem is made available to participants through the GitHub project page. The tutorial uses Python and the scikit-learn machine learning library.

Participants are encouraged to use a combination of data science techniques and engineering knowledge to come up with the best possible model.

3.2 Submitting contributions

Because the majority of data science tools and techniques have been developed as a collaborative community effort, this prediction event will also be organised as an open competition.

Contributions from participants will be uploaded to the GitHub project page to allow others to learn from modelling approaches proposed by others. Participants can refine their models multiple times and are encouraged to submit each revision.

Participants can select a team name for submitting contributions and through this team name, decide whether their contribution is anonymized or not.

Contributions can be submitted via email (bruno.stuyts@vub.be) or for participants familiar with git, through GitHub. The proposed prediction model must be submitted in a useable form to check performance against the validation dataset (e.g. Jupyter notebook).

3.3 Summary and presentation of results

All predictions will be evaluated using the validation dataset of 20 piles excluded from the data provided. A model score (e.g. median accuracy) will be used to rank the different predictions.

The lessons learnt from the prediction event and the best models will be discussed in an ISFOG2020 paper.

3.4 Timeframe

The prediction event runs for a period of 8 months, from April to December 2019. This time frame is long enough to allow teams to build up their models iteratively and learn from other submissions.

3.5 Communication and support platform

During the prediction event, there will be frequent communication to encourage contributors, provide links to machine learning tutorials, provide hints for improving models and discuss problems people might be facing.

The GitHub project page will be used for troubleshooting and support to participants who are stuck or have code-related questions. This Issues section of the GitHub project page will be used for this purpose.

Frequent announcements on the status of the prediction event will be posted on a LinkedIn page and discussion on this page is encouraged to boost the collaborative nature of this exercise.

Links to the GitHub and LinkedIn pages will be available on the ISFOG2020 website.

4. REFERENCES

Alm, T., Hamre, L., 2001. Soil model for pile driveability predictions based on CPT interpretations. Presented at the International Conference On Soil Mechanics and Foundation Engineering.

Smith, E.A.L., 1960. Pile driving analysis by the wave equation. Soil mechanics and foundations division - ASCE 24/01/1900.

Stevens, R.S., Wiltsie, E.A., Turton, T.H., 1982. Evaluating Pile Drivability for Hard Clay, Very Dense Sand, and Rock. Presented at the OTC 4205, 14th Annual Offshore Technology Conference, Houston, Texas, pp. 465–469.

5. ACKNOWLEDGEMENT

The dataset is kindly provided by Cathie Associates (www.cathie-associates.com).