

Graduation project/internship at Biodentify:

Using Machine Learning and Terabytes of Soil Microbe DNA Data to Accurately Find Oil & Gas Prior to Drilling

Important information

- Duration:** 3/6 months up to 1 year (either internship or graduation project)
- Application deadline:** September 24th 2018
- Contact:** Dr. Chris te Stroet (chris.testroet@biodentify.nl, +31 6 51246355)
- Company:** Biodentify BV, Delft
- Working location:** Delftechpark 25, Delft
- Start:** October 1st 2018 or sooner
- Fin. compensation:** 600 euro gross per month, excl. travel expenses etc.

About us

Biodentify is a company that estimates productivity of hydrocarbons at the hand of analyzing a tiny shallow soil sample on DNA. Part of the microbial ecosystem reacts on the presence of oil and gas in the deeper subsurface. This is because a very small amount of gas bubbles is moving upwards by buoyancy through micro cracks in the subsurface. This amount is too small to measure, but some of the microbes are growing on it and are more abundant above an oil and gas field. In this way the microbes form a very sensitive indirect sensor. For a 3-minute video on our technology: <https://vimeo.com/188991645> (or Google on: Vimeo and Biodentify).

Your challenge

Our workflow is summarized by: 1) sample soil, 2) DNA fingerprint analysis, 3) train a sweet spot prediction model and 4) predict the sweet spots (see Figure 1). Your challenge is to find small differences in DNA fingerprints above hydrocarbons with Machine Learning (in phases 3 and 4). In other words, to unravel the tiny differences within Terabytes of DNA fingerprint data of the ecosystem that lead to those microbes that live or die from the hydrocarbon bubbles. In 1 mm³ of soil, hundreds of thousands of species can be found; only 50-200 of them are biomarkers for hydrocarbons.

More specifically, we have 3 topics to work on:

- 1) Dimensionality reduction or feature selection. There are hundred thousands of species in the microbial ecosystem. Within their habitat, soil organisms are eating, respiring, competing, cooperating, and responding to changes in their immediate environment. The soil community and its habitat are influenced by an interconnected web of variables that differ among ecosystems, making each ecosystem somewhat unique in its microbial community. Within evolution time a huge amount of species has specialized themselves to all kind of circumstances. Before being able to use a classification algorithm we shift 'the noise from the potential signal', this means species that are being filtered out because they are not important. We want to do this without using the labels (yes/no above a hydrocarbon) and using features in the DNA fingerprints (species and how often they are count) only.

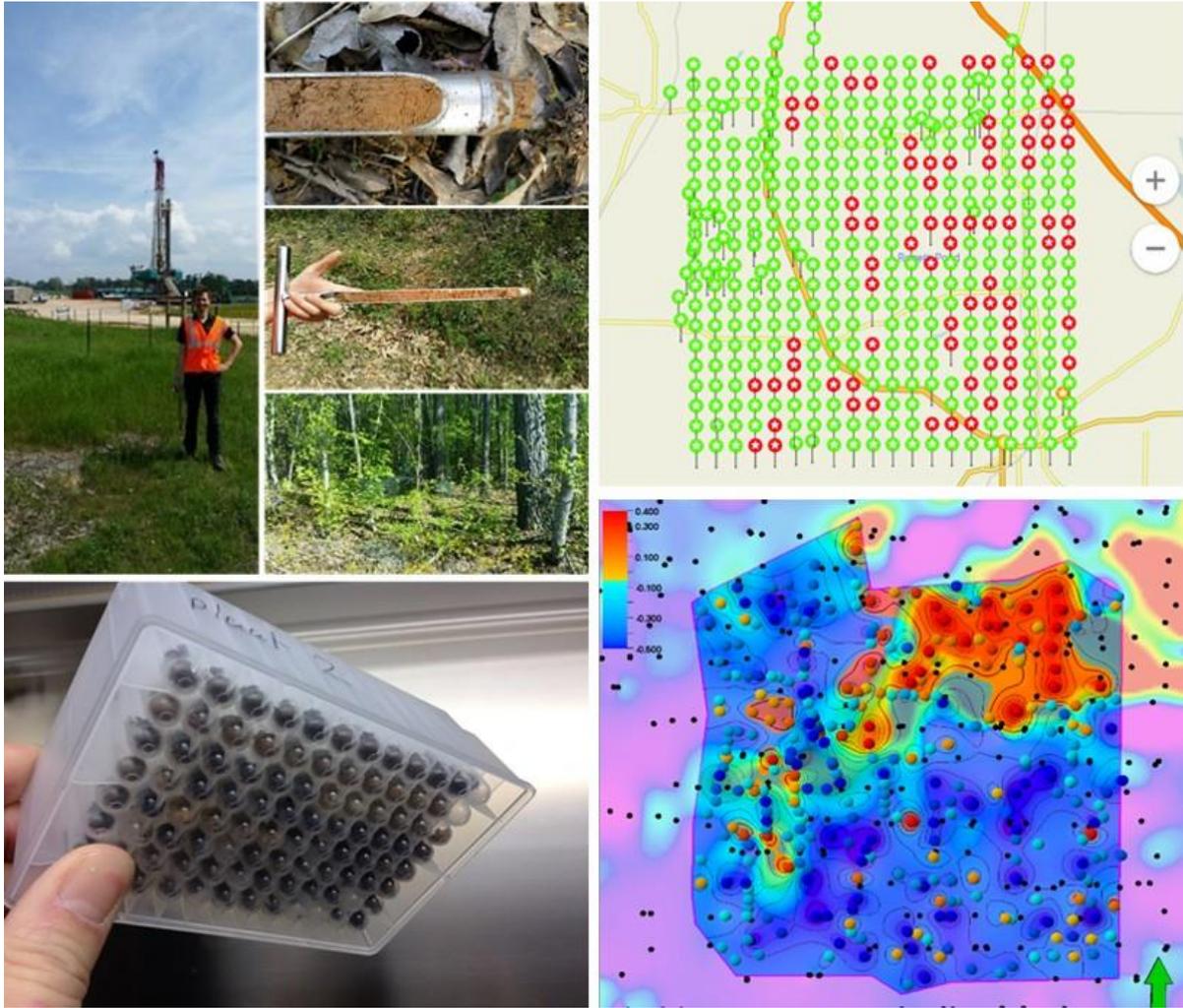


Figure 1. Our workflow is summarized by: 1) sample, 2) DNA analysis, 3) train a model and 4) predict.

- 2) Semi supervised SVM classification. After feature selection we now use an SVM classifier. We do this in a supervised algorithm; we only use the DNA fingerprints from our database with known labels (2500 samples). We would like to extend this by including the use of new DNA fingerprints where we must predict the class, so semi supervised estimation. This to increase the accuracy of the prediction.
- 3) Use deep learning algorithms for dimensionality reduction and/or classification. Because our database is growing, we would like to use the potential by implementing Deep Learning neural networks for our problem. This is virgin area for us.

Your background

We are looking for students with the following background:

- 1) Studying a Master's Degree of Artificial Intelligence or comparable
- 2) High grades for relevant courses (esp. on machine learning)
- 3) Ready for internship or final graduation (Master's) project
- 4) Preferably no other projects or courses
- 5) Familiar with big data processing (e.g. DNA or comparable data)
- 6) Smart, pro-active and team player
- 7) Fluent in written English
- 8) OK to work full-time + at least several days a week at our office

Our offer

We offer a position up to one year with good financial compensation and guidance of three top modelers/Machine Learning specialists. There is also support to travel to present the results in the international community. The position is based at Biodentify's main office in Delft, Delftechpark 25.

We are a growing company and got a large EU innovation grant for developing the company further: <https://www.rvo.nl/actueel/nieuws/9-miljoen-voor-5-nederlandse-innovatieve-scale-ups>. If you want to get experience in working for a high-tech European startup that tests the limits of current AI technology, JOIN US!

Application

Interested? Please send your **application letter**, **list of grades** and **CV** to chris.testroet@biodentify.nl before the application deadline. Any questions are also welcome. See you soon!