GAZPROM NEFT DIGITAL

OPEN TOOLS FOR ADVANCED DATA SCIENCE IN EXPLORATION

DEPARTMENT

OF MACHINE LEARNING IN GEOLOGY

GPN ML&DS EDUCATION COURSE

COURSE CONTENTS

SUMMER ONLINE 2020

Duration of daily lectures
2h30min demonstration\practice
30min discussion, elimination of critical issues
Additional: 1 hour for self-dependent studies

Day 1:

First-break picking
Analysis of regression and segmentation problems

Day 2:

Well logging curve reconstruction
Automatic core-to-log matching
Porosity prediction

Day 3:

Seismic geometry QC
Denoising seismic data

Day 4:

Seismic horizon detection
Horizon interpolation approach

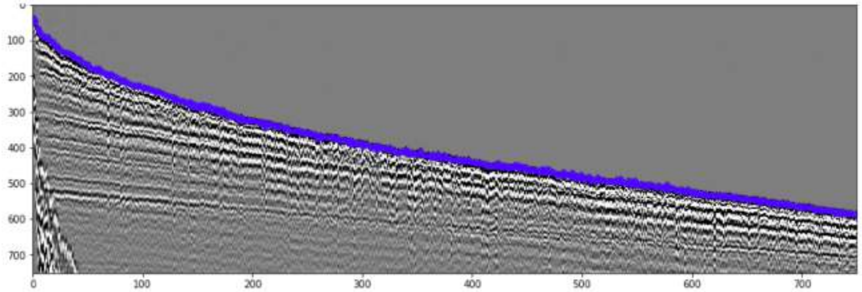
Department of Digital Transformation, LLC GPN
All materials are available in the corporate repository
https://github.com/gazprom-neft

1.
FIRST-BREAK PICKING

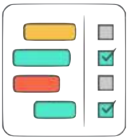
We will begin the study of the seismic processing pipeline with **first-break picking**, the results of which are important for all subsequent stages of processing and interpretation.

First, we examine the dataset containing the initial data - seismograms from a real field, and the targets - the first arrivals marked earlier by the expert.With this data we will train the model and then check how well it performs.
To do this, we will study the metrics for assessing the quality of models and determine which are suitable for this task.

The problem at hand can be solved in several ways. We will try two approaches to the problem of picking the first arrivals - regression and segmentation, and determine which one is more promising.

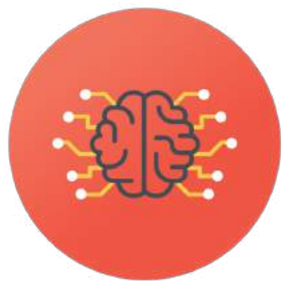


After that, it is of crucial importance to check how well the model can work on a field that is has never seen - by this means we will be able to examine the ability of the model to generalize.



Here you will have to learn to reasonably praise the model, as well as to constructively criticize - correctly find errors, investigate them and fix them by training advanced models

At the end of the lesson, you will have a full-fledged neural network model for first-break picking, which works well on real data.



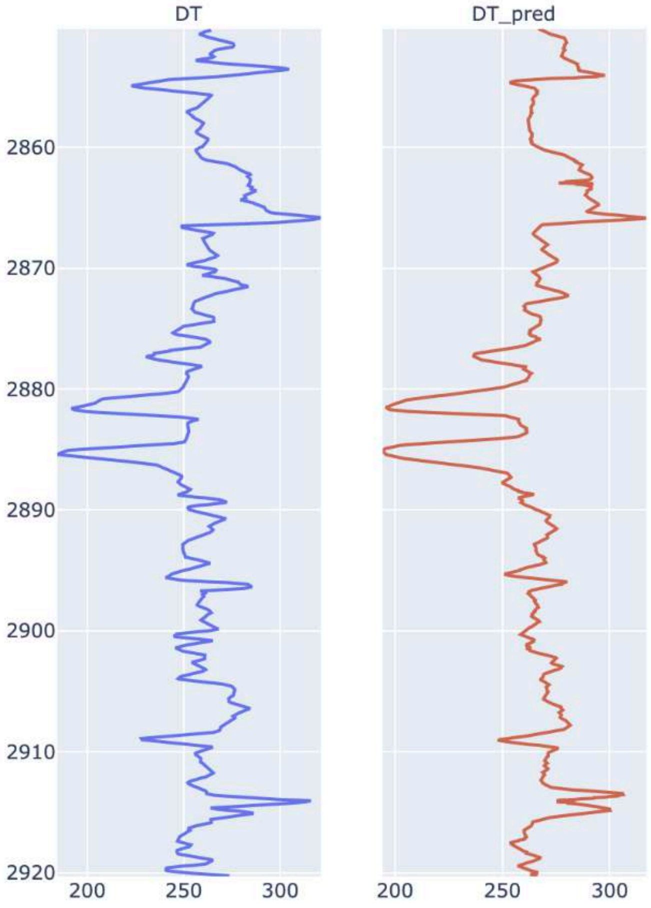
2.

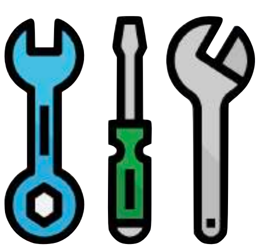
WELL LOGGING CURVE RECONSTRUCTION

Reconstruction of sonic log is a problem well-known for every petrophysicist. **Getting the logging curve** where it was not measured **is a task of great importance** (i.e. for tying well data to seismic measurements for the purposes of seismic interpretation).

For now, the most popular approach to this task among most petrophysicists is combining several linear regression models, each tied to a single stratum layer. Needless to say, the depths of the intervals are always chosen by the expert manually. These nuances make the model itself less flexible, prone to cognitive biases and impossible to automate
In addition to the above, training and validating the models on the same data is a common practice in industry, which is completely unacceptable even from the point of common sense. In data science this is called a data leak and should be avoided at all costs. We will explain why.

Solving this problem can be delivered by Encoder-Decoder type neural network — a special modification of well-known UNet architecture. It is also capable of solving the regression problem, as well as the conventional linear regression algorithm, but performs much better due to its complexity and ability to obtain local spatial information.





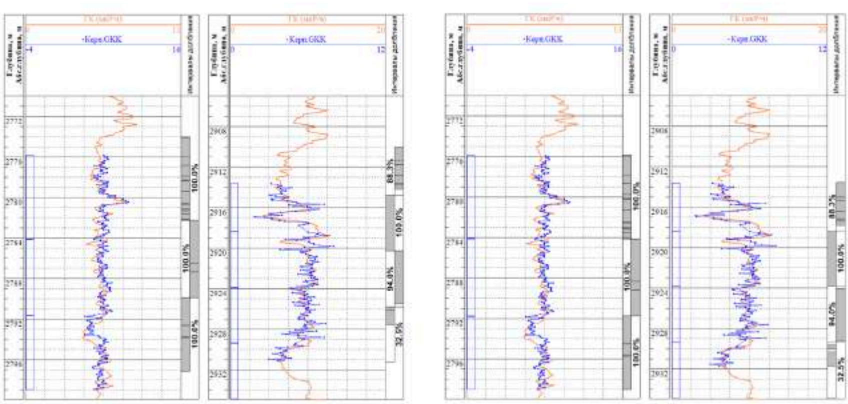
Finally, we will compare this model with the classic one, tell you how to choose the architecture and parameters, and won't forget about the preprocessing and metrics.

AUTOMATIC CORE-TO-LOG MATCHING

We will demonstrate an algorithm for core-to-log matching, which shifts coring intervals so that a well log and the corresponding core log of the same type match each other as much as possible in the sense of R2.

In the absence of logs of the core, matching can be performed by comparing a physical property of a formation with a log, theoretically correlated with it (e.g. porosity and a density log).

The proposed algorithm allows you to perform matching in just a few seconds, while manual processing takes about an hour for each exploratory well.



R^2 = 23.9 R^2 = 69.7
Expert matching

R^2 = 25.1 R^2 = 73.1
Auto matching

This task is a good example of how you can automate a routine procedure using mathematical methods.

Similar approaches have been proposed before, however, at present, many commercial packages still do not have such an opportunity.

The lesson provides details of the proposed algorithm, including its limitations, as well as demonstrates the results of its work in real time

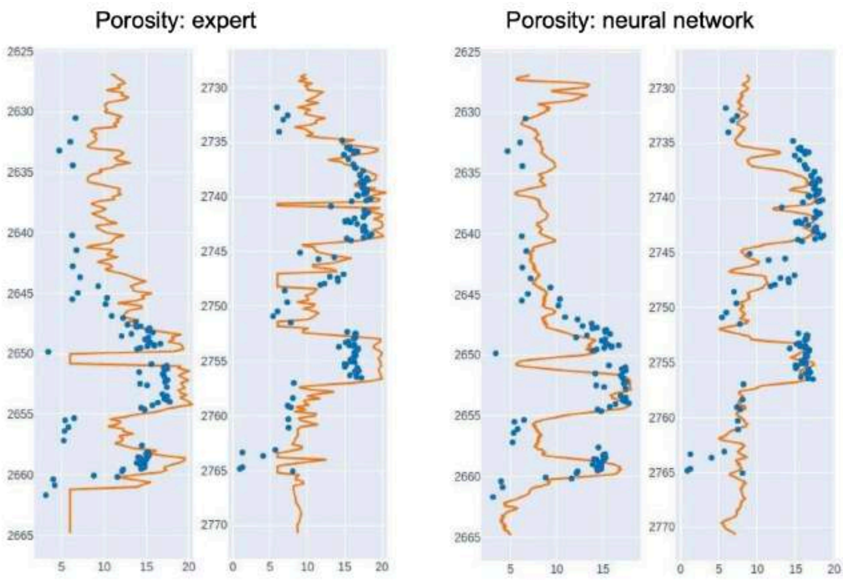
POROSITY PREDICTION

Porosity is now predicted using a set of linear regressions across the strata. Such models are not flexible enough to build an accurate prediction and do not take into account the behavior of well logs in the neighborhood of core plugs, since they are built pointwise.

Moreover, models are often trained and tested on the same data, which reduces their usefulness due to an effect that is called overfitting: the constructed model is accurate for the examples on which it was trained, but has a weak generalizing ability and shows poor quality on new data.

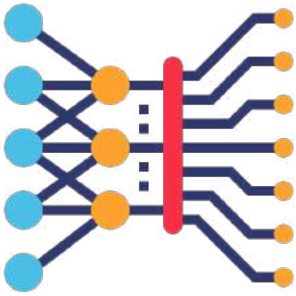
To solve this problem, we use neural networks which allow us to learn more complex dependencies from the data. Also, with this example, we will talk in detail about the correct model validation technique.

In the end, we will demonstrate a neural network for predicting porosity and compare it with expert predictions, and also talk about the shortcomings of the approach used and ways to improve it



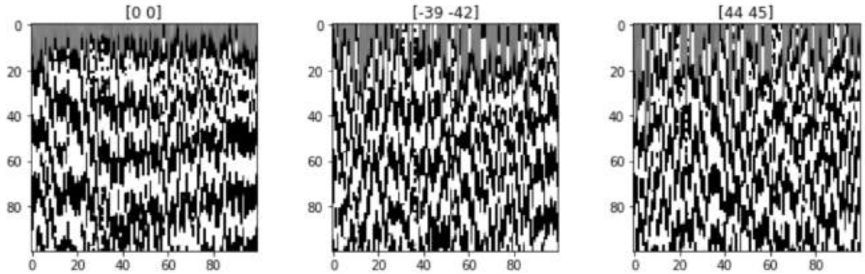
3.
SEISMIC GEOMETRY QC

The next seismic processing task is to determine the quality of seismic geometry. Studying the dataset, we will encounter the problem of a small amount of training data and discuss what approaches can be used to solve this problem.



Next, we will examine the ability of the model to generalize, and at the end of the work we will take a very important step - define a set of ideas and hypotheses that can improve the model and which need to be checked at the next stages of work.

In this class we will show you examples of how data augmentation can look and how “powerful” it can be depending on the task.



As a result, you will learn what you can do with small data and how to apply augmentation. And also you will get a ready-to-use model for quality control of geometry assignment, capable of finding displacements up to 20 meters.

DENOISING SEISMIC DATA

Next in the seismic processing unit is the noise reduction task. First, it is necessary to examine the existing dataset, consisting of seismograms before and after the noise reduction procedure.

Then we discuss and select the appropriate metrics to assess the quality of surface wave cleaning. And finally, each one participant can start to build training and validation procedures and check the quality of the model at different fields.

This training session is based on demonstration of how to train a model to perform additional tasks - in this example, the model will learn to segment ground roll area in seismic traces. Followed by a discussion in detail on why objective quality metrics are so important - not only for machine learning tasks, but also for the classic algorithms.

After this lesson, you will know how consistent are evaluations by experts and why quality metrics are so important.

4.

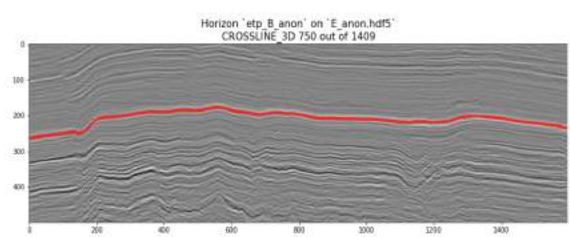
SEISMIC HORIZON DETECTION

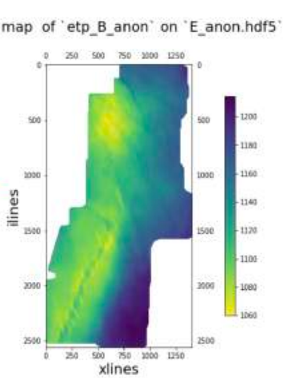
Horizon detection is a clear example of a task, where one cannot go without concise and well-defined metrics: it is completely impossible for a human to assess the quality of a tracked reflection along enormous amounts of seismic data. To this end, we have developed an open AI tool that can evaluate the horizon without human help, and share our insights and examples of its application.

Using this task as an example, we will reiterate importance of all the elements of the ML pipeline: defining the task, dataset, models and the learning process, validating with metrics and plans for further development.

Later on, details will be given on how various partitions of a dataset into train/test correspond to a completely different problems and results.

You will see how a priori expert knowledge about underground structures can be used to create geologically significant augmentations.





As in most of our other tasks, segmentation neural networks are implemented (encoder-decoder architectures) for training on 3D crops cut from a seismic cube.

Using a set of horizons as an example makes it clear that without computable metrics that it is completely impossible to build usable models: assessing the quality of horizon tracking over the entire area is very difficult even to experts.

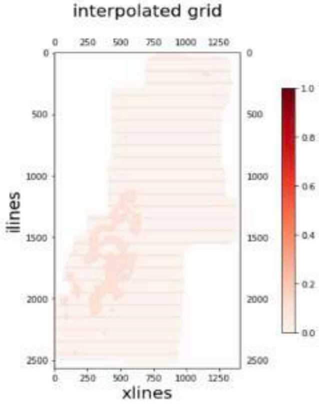
We have developed an open AI tool that can evaluate the horizon without human help, and you will be actively involved in demo-examples of its application.

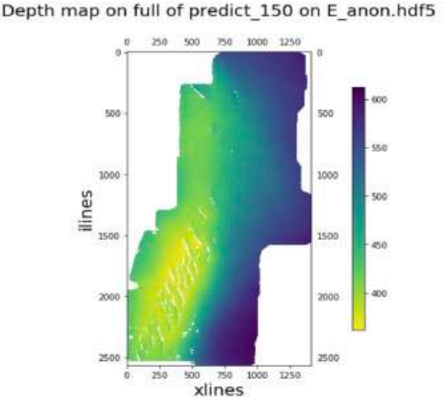
HORIZON INTERPOLATION APPROACH

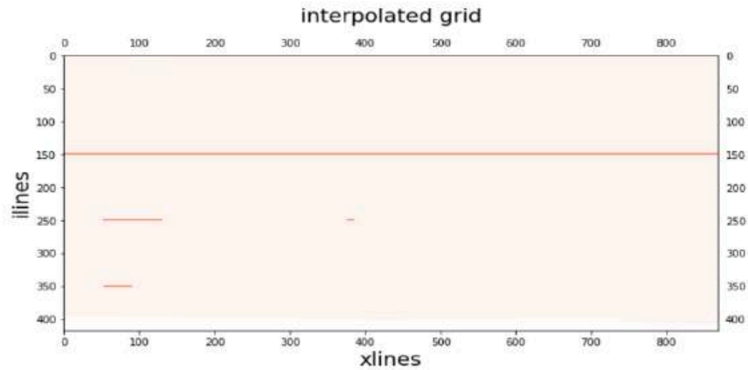
In this approach, we use a small part of the cube (<3%) for training with the goal of marking up the entire remaining cube. For example, an autocorrelator fits into the same setting!

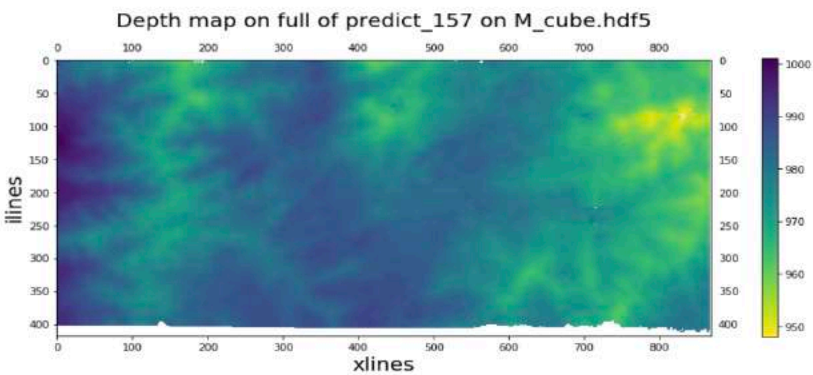
In this lessons metrics that we developed earlier are used to build a map of the complexity of the cube and calculate the training grid, then train the model in real time and look at the finished horizons. After that we will compare them with the results of the autocorrelator from Petrel.

Naturally, the results obtained by the model depend on the frequency of the grid, so we are exploring different options for the discharged grid for training.

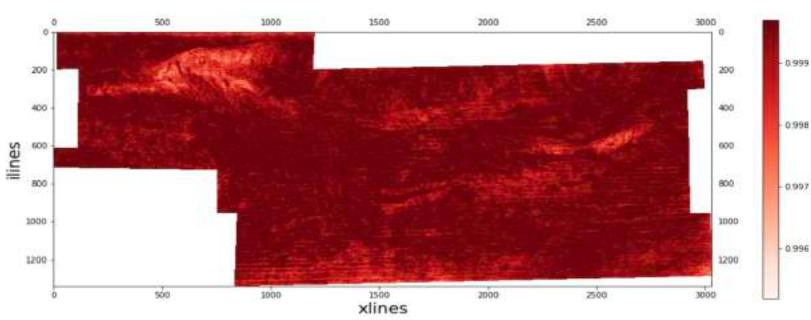






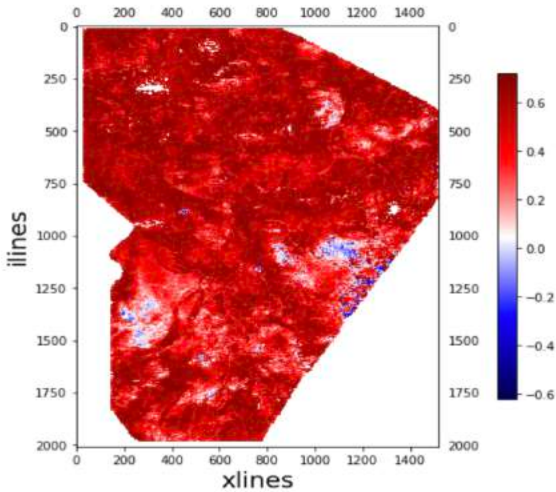


On the other hand, different sets of seismic cubes can be used for training and validation: we study on one cube, with inference on the others



In this approach, one cannot do without data normalization, as well as without augmentations that allow one to correctly take into account the uniqueness of data within each individual geology.

As a parameter for the study, we look at the saturation of the model with data: with an increase in the number of cubes in the training set, the quality of predictions increases.



Combining both approaches, we can improve the results of each of them: we will also look into technical details of the implementation of such combination. In addition to the task of tracking horizons, we will take a look at the possibility of distinguishing other geological bodies, such as seismic facies and river estuaries.

SEE YOU IN CLASS!