

How-to use:



This document is a step-by-step tutorial on running the WellCast web-app using the provided sample data.

STEP 1. Download sample file

We provide a dataset that could be used in this tutorial.

The screenshot displays the WellCast web application interface. At the top, the WellCast logo is visible. Below the header, a text block explains the application's purpose: predicting missing well log data using a machine learning workflow based on the SPE GCS ML Challenge 2021. A 'Workflow Diagram' section shows a flowchart of the process: 1. Import .LAS Files, 2. Select training data, 3. Correlation heatmap assessment, 4. Select target log, 5. Visualize using training data, 6. Model creation, 7. RMSE and feature importance, 8. Introduce new well, and 9. Download result. A decision diamond labeled 'Satisfied?' leads to 'Re-model' if 'No' or 'Introduce new well' if 'Yes'. The 'Input File' section at the bottom includes a 'Choose Files' button, a 'Submit' button, and a 'Download Sample' button. A blue arrow points to the 'Download Sample' button with the label 'Click'.

WellCast
GEO VARTHA

WellCast is a web-application tool for predicting missing well log data using the machine learning workflow. In this application, we used gradient boost method (tree-based algorithm) which was adapted from the result of [SPE GCS ML Challenge 2021](#). This workflow summarized on the picture below.

Workflow Diagram

WORKFLOW DIAGRAM

1. Import .LAS Files
2. Select training data
3. Correlation heatmap assessment
4. Select target log
5. Visualize using training data
6. Model creation
7. RMSE and feature importance
8. Introduce new well
9. Download result

1. Upload .las files
2. Las files selection
3. Exploratory Data Analysis
4. Model Building
5. Prediction result

3. Select logs that will be used as predictor and target

4. Filter logs

5. Satisfied?

Re-model

1. Input File

Select .las file to be uploaded

Choose Files No file chosen Submit

All las files uploaded will be removed after the session is closed or after 60 minutes.

Click

The zip file contains 4 train wells and a new well on (.LAS) format. We will predict a log from the new well file at the end of this tutorial.

STEP 2. Upload files and preview

1. Input File

Download Sample

Select .las file to be uploaded

Choose Files

4 files

Submit

All las files uploaded will be removed after the session is closed or after 60 minutes.

train1.las

train2.las

train3.las

train4.las

Upload the four train wells and leave the new well for now. We need to check the .LAS files before merging them as the model creation dataset. Please make sure all the logs' mnemonics on your wells are listed on this web-app dictionary, rename them temporarily if necessary.

1. Input File

Download Sample

Select .las file to be uploaded

Choose Files

4 files

Submit

All las files uploaded will be removed after the session is closed or after 60 minutes.

100%

2. Select .las files for training

Click on the preview button to view the logs of the uploaded las files.
Subsequent to the visual inspection, select the las files that will be used for the training process.

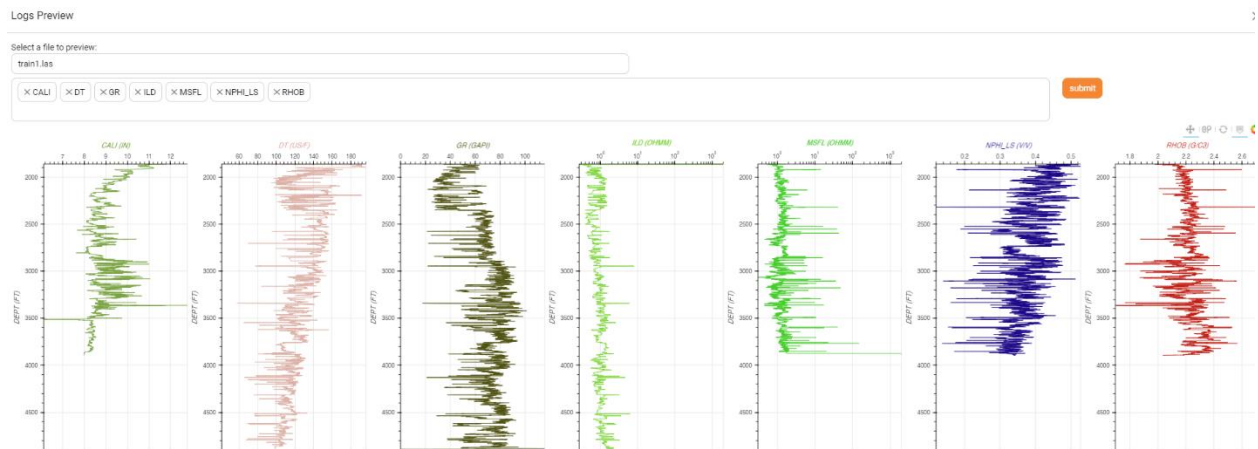
.las files:

Click here to select files

Preview Logs

Submit

Disclaimer: uploaded logs file will be renamed, and existing log header will be aliased using our library. Click [HERE](#) to see the aliasing library.



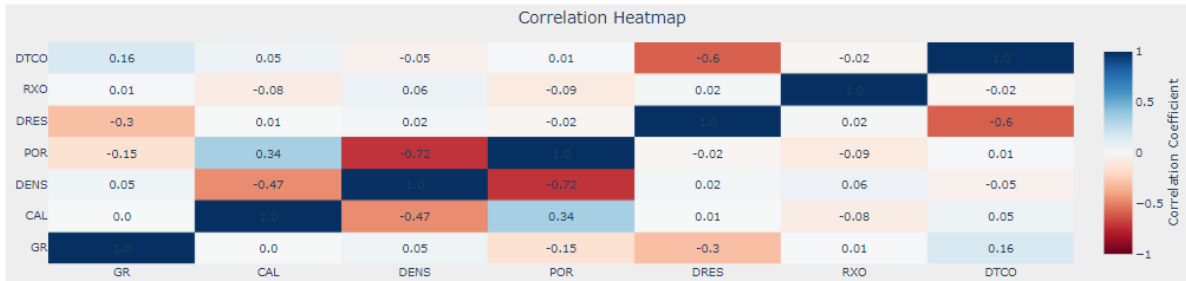
STEP 3. Data exploration

In this web-app we simplify the EDA phase which is usually far more complicated than the actual cases. However, these minimal EDA steps were proven to be efficient on our latest project at SPE GCS ML Challenge 2021. Feature selection based on Pearson's correlation is performed in the first step of EDA.

3. Exploratory Data Analysis (EDA)

Calculate the correlation coefficient between all the different logs included in the data^[1].
Colorbar shows correlation coefficient value.

High value (1) shows positive correlation, low value (-1) shows negative correlation, while 0 shows no correlation



Based on the correlation heat map, please choose the log data to be included for the model training (including the target log)

☐ GR ☐ CAL ☐ DENS ☐ DRES ☐ DTCO

GR

CAL

DENS

POR

DRES

RXO

DTCO

Then, remove the possible outlier on each log. We can define the minimum and maximum value based on our domain knowledge or based on statistical perspective as shown in this tutorial. In this example, we found an outlier on our DRES data, so we cut the maximum DRES as the q3 from the box plot. Click preview to display the new boxplot and submit to proceed.

i Based on the correlation heat map, please choose the log data to be included for the model training (including the target log)

Submit

i You can remove potential data outlier by modifying the minimum and maximum value accepted for each logs.

Log	Minimum	Maximum
GR	0.00	115.69
CAL	5.85	22.87
DENS	1.32	2.72
DRES	0.27	1.091
DTCO	57.63	264.00

Before

After

i Make sure that you are satisfied with the data cleansing process, we will create a model based on this data by clicking submit button.

STEP 4. Model Building

For this Beta version, we will use a gradient boosting method to predict our missing log. We split the cleansed data from the previous step into 80% of training set and 20% of validation set. In this web-app, we only need to define which log that we would like to predict. In this tutorial we will predict the DTCO log.

4. Model Building

The supervised learning method deployed here is a gradient boosting method[\[2\]](#).

The data samples from the training wells are divided into training and validation samples with the proportion of 80 and 20 percent respectively.

Select log you want to predict

DTCO

Submit

Assess the Root Mean Square Error (RMSE)[\[2\]](#) of the train & validation samples.

RMSE train	8.109
RMSE validation	12.368

Assess the predictor parameters based on the feature importance

Log	Feature Importance
GR	0.08
CAL	0.05
DENS	0.15
DRES	0.70

Select file to be used for prediction:

Choose File No file chosen

Submit

You can assess the model quality by assessing the RMSE on train and validation dataset. You can also check the result visually by using the model to predict the train dataset. If the result is not satisfying, you can repeat the workflow. Maybe introduce more wells or the same dataset but different log selection.

Select file to be used for prediction:

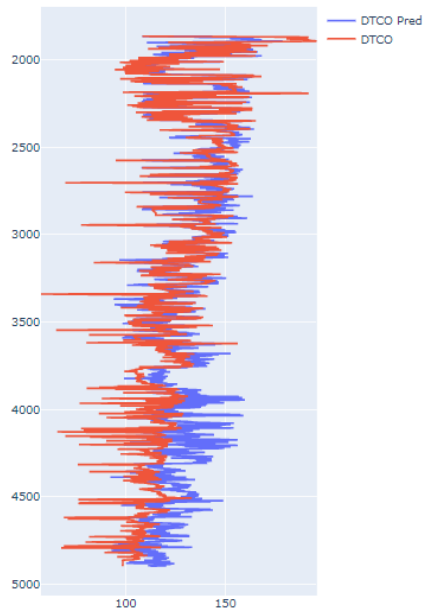
Choose File train1.las

Submit

5. Prediction result

Plot the predicted log with the actual log

Open Uploaded file logs



Disclaimer: this workflow might not produce a good result on your dataset. The aim of this web-application is to facilitate a non-programmer geoscientists who are willing to learn and try the implementation on machine learning workflow on missing well log prediction.

Download

STEP 5. Prediction result.

If we are satisfied with the model, we can use it to predict the DTCO log from a new well. Replace the train file with the new .LAS file on the previous step.

Select file to be used for prediction:

Choose File

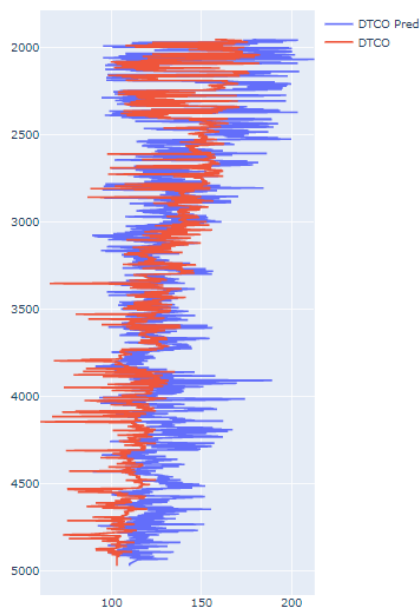
new_well.las

Submit

5. Prediction result

Plot the predicted log with the actual log

Open Uploaded file logs



Disclaimer: this workflow might not produce a good result on your dataset. The aim of this web-application is to facilitate a non-programmer geoscientists who are willing to learn and try the implementation on machine learning workflow on missing well log prediction.

Download

Disclaimer: this workflow might not produce a good result on your dataset. However, there is a good chance that you will get a fair result as we got acceptable output from several different dataset using this workflow. The aim of this web-application is to facilitate non-programmer geoscientists who are willing to learn and try the implementation of machine learning workflow on missing well log prediction. In the future development of the web-app we will add other methods and parameter setting features to give users a better experience on the machine learning experiments using well log data.

References:

1. Sedgwick, Philip (2012). [Pearson's correlation coefficient](#)
2. Ayyadevara, V Kishore (2018). [Gradient Boosting Machine](#)
3. Chai, T., Draxler, R. R. (2014). [Root mean square error \(RMSE\) or mean absolute error \(MAE\)?](#)
4. Doa Ibu Team (2021). [SPE GCS ML Challenge 2021](#)