ML Competition - Datathon

Out date: June 09, 2022 (Thursday) 9:15 AM – 11:15 AM

Submission

1. Prepare your solution in Orange and save the workspace as ows (TeamName.ows).
2. Deploy your models to predict the target variable in the test dataset.
3. Submit predictions to the mock Kaggle competition: https://www.kaggle.com/t/ea01fb40475c4405a8b8b33700aba528
4. Upload your ows file to https://forms.gle/6Bw4dm2gfNgnkqPw5

Objective: Your customer is asking you to use Machine Learning to predict **Facies** in two wells (Well-A and Well-B), contained in the *topredict_facies.xlsx* file. As a data scientist, your job is to train machine learning models using the *Train.xlsx* file provided to you using **Orange,** deploy your best performing model to predict **Facies** for the two wells and submit the predictions to your client.

Data: Please download *Training.xlsx* and *topredict_faciles.xlsx* from Kaggle to your laptop.

Seven predictor variables:

Five wire line log curves include gamma ray (GR), resistivity (ILD_log10), photoelectric effect (PE), neutron-density porosity difference (DeltaPHI), and average neutron-density porosity (PHIND). Log Depth for each well is also provided.

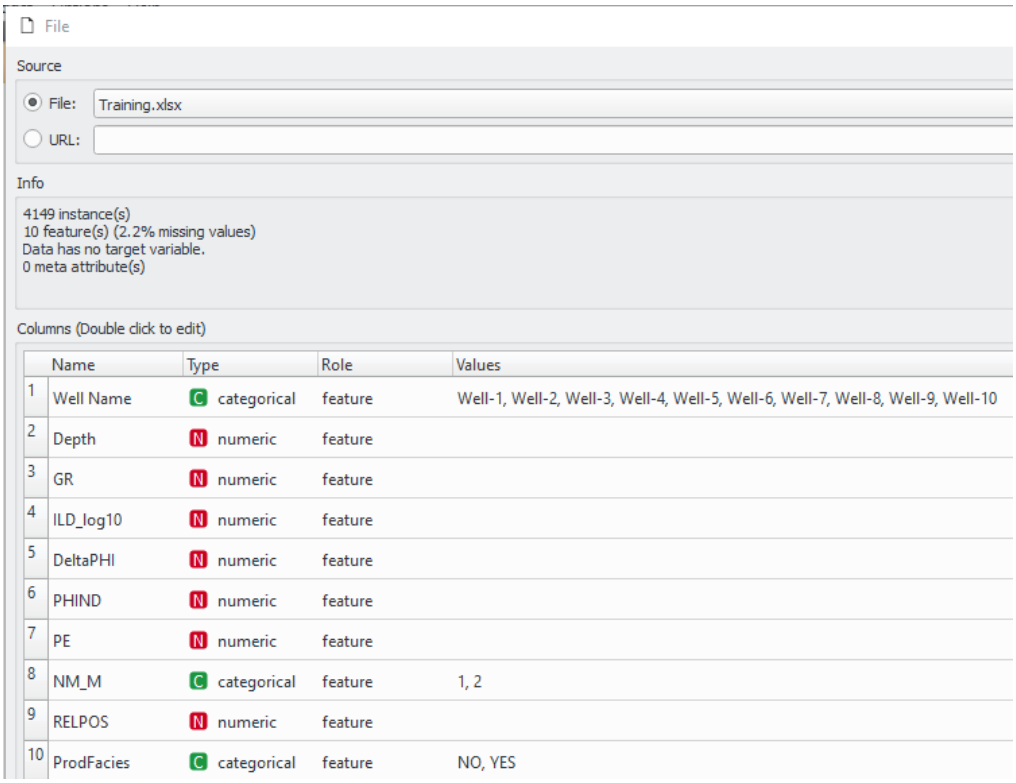
Two geologic constraining variables: nonmarine-marine indicator (NM_M) and relative position (RELPOS)

Target variable: Facies, 2 classes- YES indicating productive facies and NO indicating non-productive facies.

Reference: https://library.seg.org/doi/full/10.1190/tle35100906.1 (dataset)

https://www.slb.com/resource-library/oilfield-review/defining-series/defining-logging

1. *Train.xlsx* containing the following features:


File

Source

File: Training.xlsx

URL:

Info

4149 instance(s)
10 feature(s) (2.2% missing values)
Data has no target variable.
0 meta attribute(s)

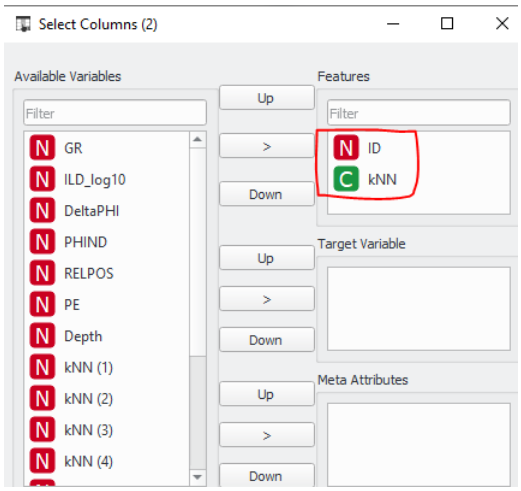
Columns (Double click to edit)

	Name	Type	Role	Values
1	Well Name	categorical	feature	Well-1, Well-2, Well-3, Well-4, Well-5, Well-6, Well-7, Well-8, Well-9, Well-10
2	Depth	numeric	feature	
3	GR	numeric	feature	
4	ILD_log10	numeric	feature	
5	DeltaPHI	numeric	feature	
6	PHIND	numeric	feature	
7	PE	numeric	feature	
8	NM_M	categorical	feature	1, 2
9	RELPOS	numeric	feature	
10	ProdFacies	categorical	feature	NO, YES

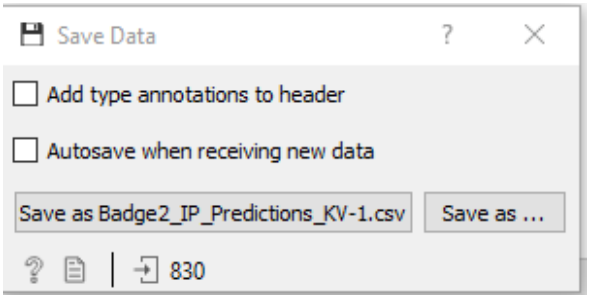
2. *topredict_facies.xlsx* which contains the same features as the *Train.xlsx* file shown above except the **ID** and **Facies** target variable.

Instructions

1. Train your models using Orange software and by loading the data from *Train.xlsx*
2. Build your Orange pipeline considering the key steps as shown during yesterday's Orange training session.
3. You are allowed to use only KNN, Decision Tree and Random Forest algorithms.
4. Once you have trained your models and ready to deploy them to predict facies for the two wells A and B, load the *topredict_facies.xlsx* file to Orange and predict facies using your desired models and the Predictions widget.
5. Use **Select Columns** widget to select only the ID and predicted facies values columns.



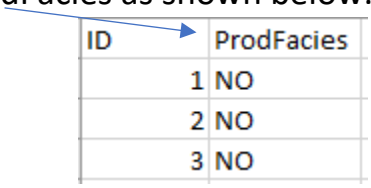
6. Use **Save Data** widget to save your predictions as csv file to your laptop. Uncheck **Add type annotations to header** and save the .csv file to your local folder.



7. Predictions csv file should output results in two columns: 1) ID, and 2) the name of ML algorithm (e.g., kNN here) as shown below.

	A	B
1	ID	kNN
2	1	1
3	2	1
4	3	1
5	4	1
6	5	1
7	6	1
8	7	1
9	8	1
10	9	1

8. Rename column B to ProdFacies as shown below.



ID	ProdFacies
1	NO
2	NO
3	NO

Note: If Kaggle gives you an error during submission, change the ID column to Text format and submit your results.

9. Upload the csv file to Kaggle: https://www.kaggle.com/t/ea01fb40475c4405a8b8b33700aba528

10. Upload your ows file to https://forms.gle/6Bw4dm2gfNgnkqPw5