1. View the raw input data

View an excerpt of the raw input data in a suited visualization. For example the first rows of a tabular data set rendered as a table.

Angola	2020-07-06	Alpha	0	0.0
Angola	2020-07-06	B.1.1.277	0	0.0
Angola	2020-07-06	B.1.1.302	0	0.0
Angola	2020-07-06	B.1.1.519	0	0.0
Angola	2020-07-06	B.1.160	0	0.0
Angola	2020-07-06	B.1.177	0	0.0
Angola	2020-07-06	B.1.221	0	0.0
Angola	2020-07-06	B.1.258	0	0.0
Angola	2020-07-06	B.1.367	0	0.0
Angola	2020-07-06	B.1.620	0	0.0
Angola	2020-07-06	Beta	0	0.0
Angola	2020-07-06	Delta	0	0.0

2. View the meanings columns in the raw input data

Get a textual description explaining each feature in the data set.

The data

- location- this is the country for which the variants information is provided;
- date - date for the data entry;
- variant - this is the variant corresponding to this data entry;
- num_sequences - the number of sequences processed (for the country, variant and date);
- perc_sequences - percentage of sequences from the total number of sequences (for the country, variant and date);
- num*sequences*total - total number of sequences (for the country, variant and date);

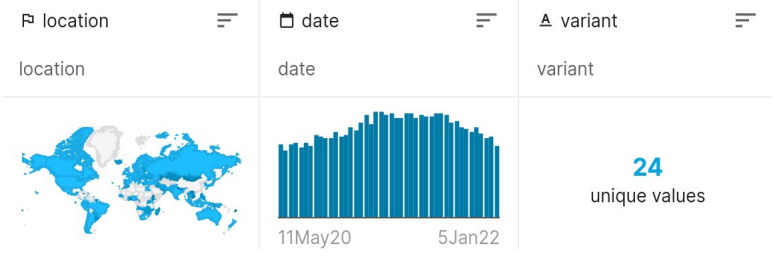
3. View statistics of raw input data

Get an aggregated view of the raw input data. The aggregated view contains basic statistics about the data. For example, a distribution of the values of a feature.

	Subject_1_Marks	Subject_2_Marks	Subject_3_Marks	Names
count	5.00000	5.000000	5.000000	5
unique	NaN	NaN	NaN	4
top	NaN	NaN	NaN	Saksham
freq	NaN	NaN	NaN	2
mean	26.80000	33.800000	26.200000	NaN
std	15.64289	12.557866	13.989282	NaN
min	12.00000	21.000000	8.000000	NaN
25%	14.00000	23.000000	21.000000	NaN
50%	21.00000	32.000000	23.000000	NaN
75%	42.00000	43.000000	34.000000	NaN
max	45.00000	50.000000	45.000000	NaN

4. Visualize the raw input data

Get an aggregated view of the raw input data. The aggregated view contains basic visualizations about the data. For example simple line plots of single features or distributions of values as a histogram.



5. View the pre-processed data

View an excerpt of the pre-processed data in a suited visualization. For example the first rows of a tabular data set rendered as a table.

Angola	2020-07-06	Alpha	0	0.0
Angola	2020-07-06	B.1.1.277	0	0.0
Angola	2020-07-06	B.1.1.302	0	0.0
Angola	2020-07-06	B.1.1.519	0	0.0
Angola	2020-07-06	B.1.160	0	0.0
Angola	2020-07-06	B.1.177	0	0.0
Angola	2020-07-06	B.1.221	0	0.0
Angola	2020-07-06	B.1.258	0	0.0
Angola	2020-07-06	B.1.367	0	0.0
Angola	2020-07-06	B.1.620	0	0.0
Angola	2020-07-06	Beta	0	0.0
Angola	2020-07-06	Delta	0	0.0

6. Know how raw input data was pre-processed

Get a textual or visual summary of all applied pre-processing operations. For example, if an encoding of categorical values was performed or if an imputation of missing values was applied.

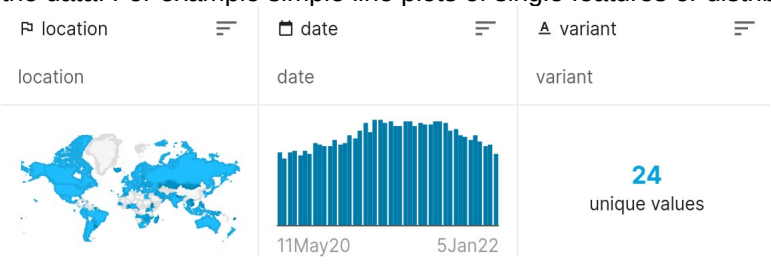
7. View statistics of data after pre-processing

Get an aggregated view of the data after pre-processing. The aggregated view contains basic statistics about the data. For example, the distribution of the values of a feature.

	Subject_1_Marks	Subject_2_Marks	Subject_3_Marks	Names
count	5.00000	5.000000	5.000000	5
unique	NaN	NaN	NaN	4
top	NaN	NaN	NaN	Saksham
freq	NaN	NaN	NaN	2
mean	26.80000	33.800000	26.200000	NaN
std	15.64289	12.557866	13.989282	NaN
min	12.00000	21.000000	8.000000	NaN
25%	14.00000	23.000000	21.000000	NaN
50%	21.00000	32.000000	23.000000	NaN
75%	42.00000	43.000000	34.000000	NaN
max	45.00000	50.000000	45.000000	NaN

8. Visualize data after pre-processing

Get an aggregated view of the pre-processed data. The aggregated view contains basic visualizations about the data. For example simple line plots of single features or distributions of values as a histogram.



9. View new engineered features

View an excerpt of the data after generating new features and selecting suited features in a suited visualization. For example the first rows of a tabular data set rendered as a table.

Angola	2020-07-06	Alpha	0	0.0
Angola	2020-07-06	B.1.1.277	0	0.0
Angola	2020-07-06	B.1.1.302	0	0.0
Angola	2020-07-06	B.1.1.519	0	0.0
Angola	2020-07-06	B.1.160	0	0.0
Angola	2020-07-06	B.1.177	0	0.0
Angola	2020-07-06	B.1.221	0	0.0
Angola	2020-07-06	B.1.258	0	0.0
Angola	2020-07-06	B.1.367	0	0.0
Angola	2020-07-06	B.1.620	0	0.0
Angola	2020-07-06	Beta	0	0.0
Angola	2020-07-06	Delta	0	0.0

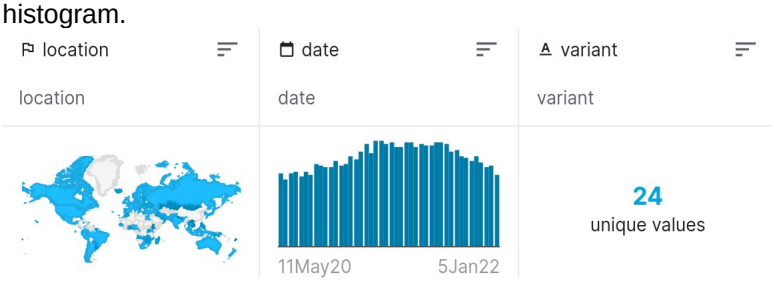
10. View statistics of data after feature-engineering

Get an aggregated view of the data after feature-engineering. The aggregated view contains basic statistics about the data. For example, the distribution of the values of a feature.

	Subject_1_Marks	Subject_2_Marks	Subject_3_Marks	Names
count	5.00000	5.000000	5.000000	5
unique	NaN	NaN	NaN	4
top	NaN	NaN	NaN	Saksham
freq	NaN	NaN	NaN	2
mean	26.80000	33.800000	26.200000	NaN
std	15.64289	12.557866	13.989282	NaN
min	12.00000	21.000000	8.000000	NaN
25%	14.00000	23.000000	21.000000	NaN
50%	21.00000	32.000000	23.000000	NaN
75%	42.00000	43.000000	34.000000	NaN
max	45.00000	50.000000	45.000000	NaN

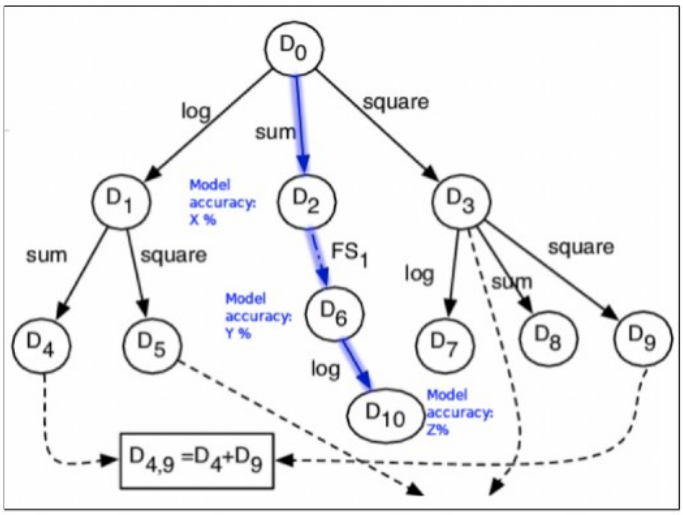
11. Visualize data after feature-engineering

Get an aggregated view of the data after feature-engineering. The aggregated view contains basic visualizations about the data. For example simple line plots of single features or distributions of values as a histogram.



12. View how new features are derived from existing features

Get a textual or visual representation how exactly existing features were used to derive new features.

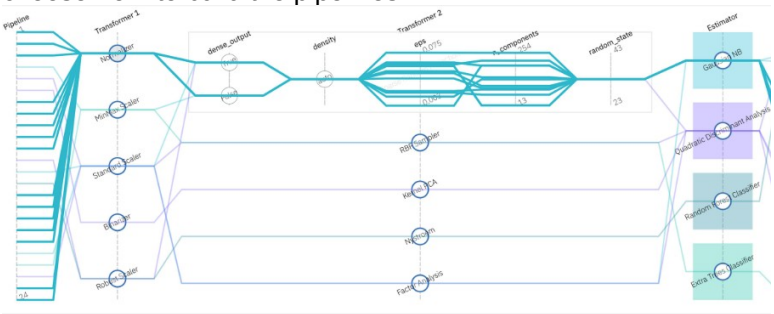


13. View the complete processing pipeline

Get a visual representation of the complete ML pipeline. This visualization contains all steps in the pipeline and shows how data flows between the different stages.

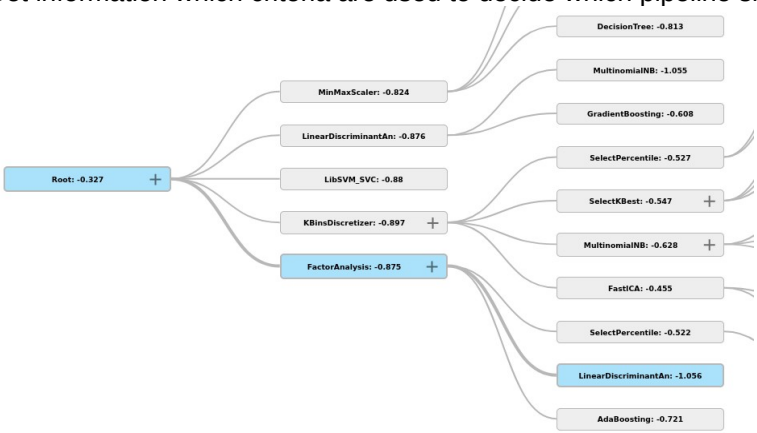
14. View which pre-processing, feature engineering and modeling algorithms are available

Know what the set of possible pipeline candidates look like, i.e., which base algorithms are available to choose from to build the pipelines.



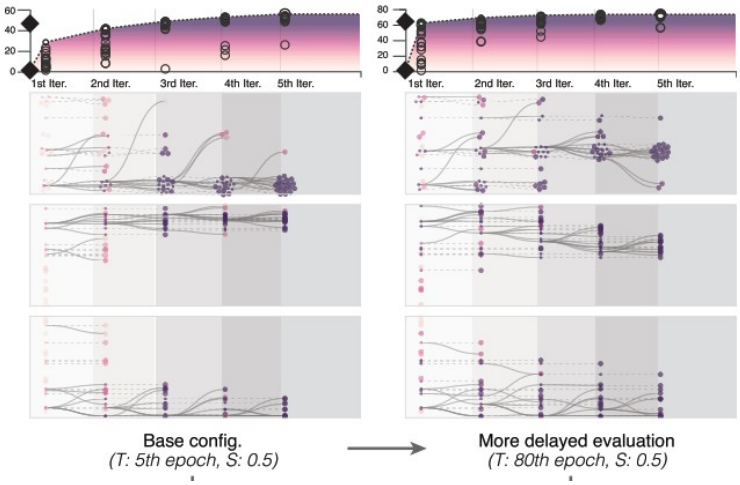
15. Know how pipelines are chosen

Get information which criteria are used to decide which pipeline should be sampled next.



16. Know how hyperparameters are chosen

Get information allowing the deduction of the internal mechanisms used to sample new hyperparameters.



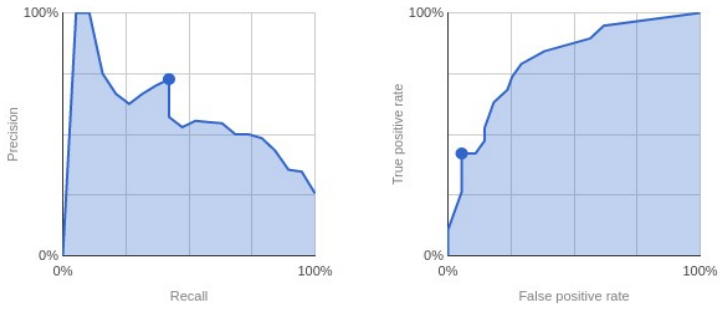
17. View evaluation metrics

Get a textual representation of various performance metrics.

F1 score	0.533
Accuracy	81.1% (60/74)
Precision	72.7% (8/11)
True positive rate (Recall)	42.1% (8/19)
False positive rate	0.055 (3/55)

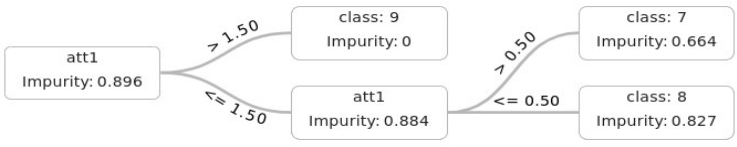
18. View visualization of evaluation metrics

Get a visual representation of various performance metrics.



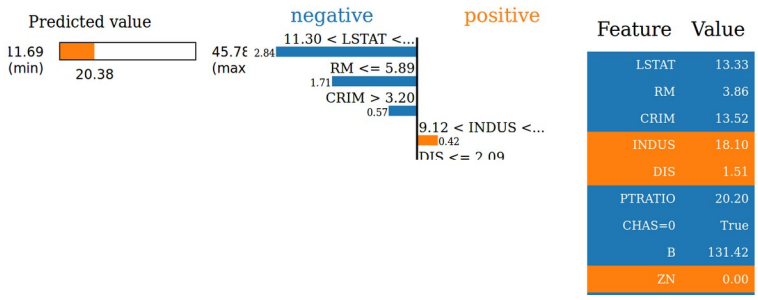
19. View global explanations for model

Get an approximate representation of the model that is interpretable. For example, replace the original model with a decision tree that behaves similar to the original model.



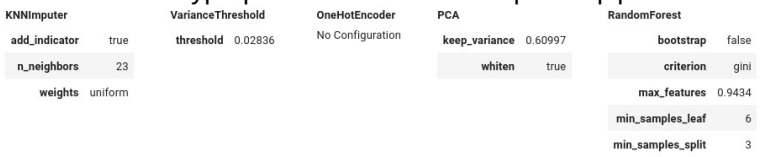
20. View local explanations for model

Get an explanation why a single sample was predicted as a specific class. For example, in form of a feature attribution to the prediction.



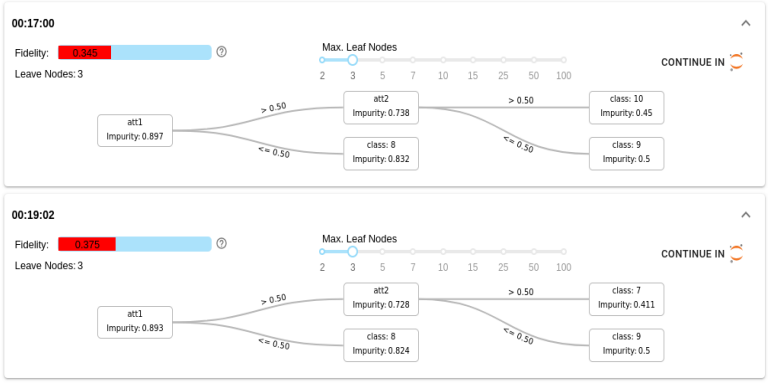
21. View hyperparameters of model

View the exact hyperparameters used in a specific pipeline.



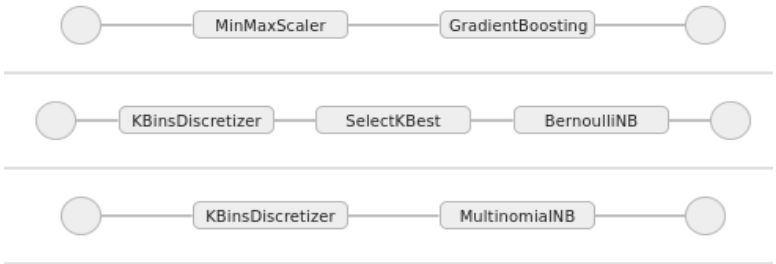
22. Compare evaluation and explanations of various model with each other

Have the option to compare the performance evaluations and model explanations of multiple pipeline candidates.



23. Compare differences between pipelines

Get a visual comparison that allows the direct comparison of different pipeline structures with the option to spot similarities/differences.



24. Compare hyperparameters of identical pipelines

Get a visual or textual representation that allows the direct comparison of the hyperparameter values of multiple pipeline candidates.

00:38:02

pca:keep_variance	0.60997	**pca:whiten**	true
preprocessing:numerical:numerical:knn_imputer:add_indicator	true	**preprocessing:numerical:numerical:knn_imputer:n_neighbors**	23
preprocessing:numerical:numerical:knn_imputer:weights	uniform	**preprocessing:numerical:numerical:variance_threshold:threshold**	0.02836
random_forest:bootstrap	false	**random_forest:criterion**	gini
random_forest:max_features	0.9434	**random_forest:min_samples_leaf**	6
random_forest:min_samples_split	3		

00:37:04

pca:keep_variance	0.95703	**pca:whiten**	true
preprocessing:numerical:numerical:knn_imputer:add_indicator	true	**preprocessing:numerical:numerical:knn_imputer:n_neighbors**	23
preprocessing:numerical:numerical:knn_imputer:weights	uniform	**preprocessing:numerical:numerical:variance_threshold:threshold**	0.01254
random_forest:bootstrap	false	**random_forest:criterion**	gini
random_forest:max_features	0.96254	**random_forest:min_samples_leaf**	6
random_forest:min_samples_split	2		