ML Process Overview with scipy-modeler

Retrieve data	Preprocess/Combine Data	Transform data	Segregate Data	Model Training & Evaluation	Select/Persist Best Estimator

Data Source 1

Data Source 2

Data Source 3

Preprocess

Preprocess

Preprocess

Merge (All data before transformations)

Column Transformer Pipeline

Final transformed data (before modelling)

All transformed data

Training/ Validation Data

Hold out (Test data)

Subsample

Training Set

Validation Set

Select best candidate model(s)

Evaluate on hold out

Persist best model for automatic inference

Using scipy-modeler API

Using scipy-modeler SKTransformer

Using scipy-modeler.model files
n holdout = n validation

Using scipy-modeler Eval

Send preprocessing pipeline to scipy-modeler manager

Create transformation pipeline with preprocessed data

Segregate data for modelling

Send back best results

Choose candidate models based on validation results

Persist model which performs best on test set

Pipeline Manager
Using scipy-modeler PipelineManager

Ongoing Model Monitoring

Design by: Nick Hopewell