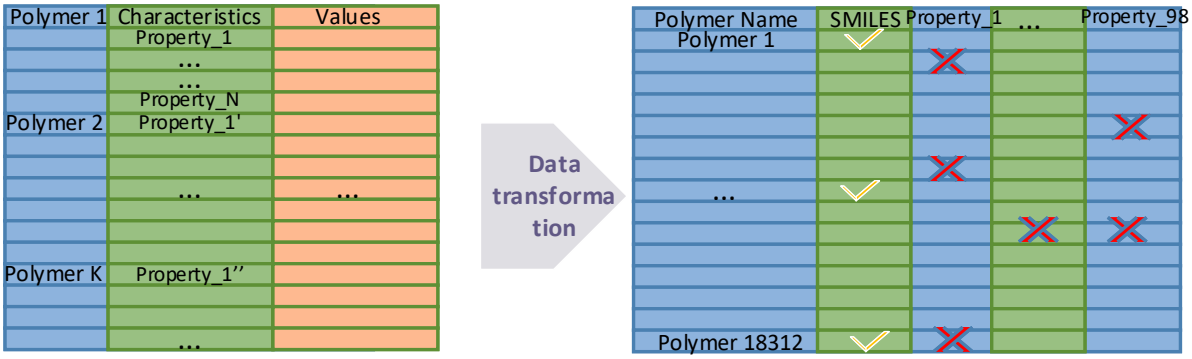
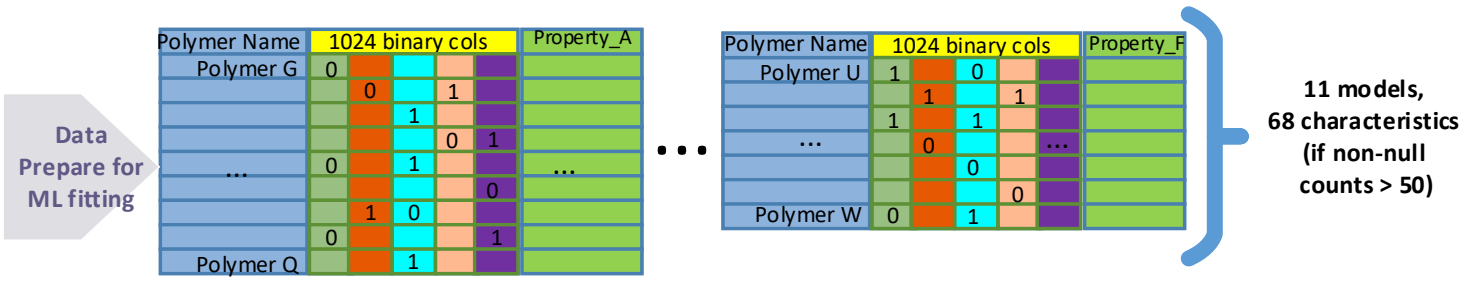
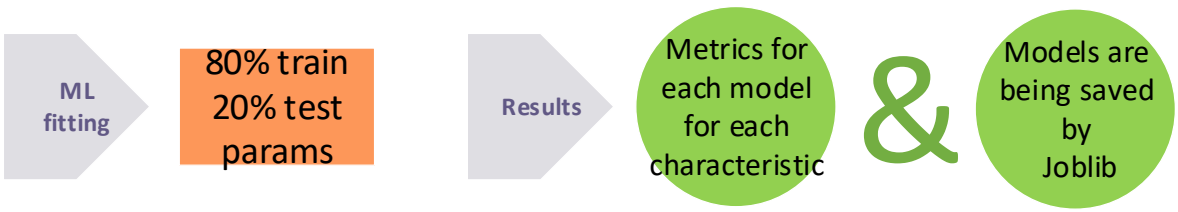
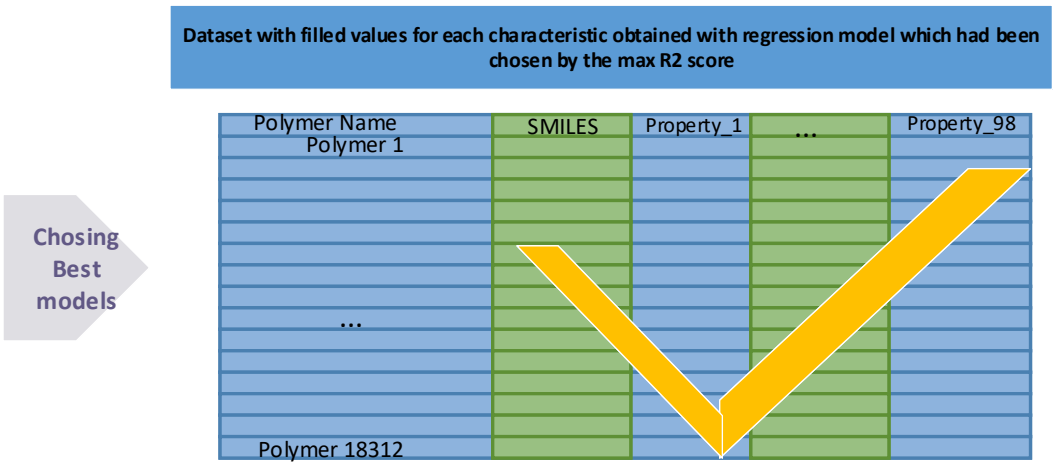
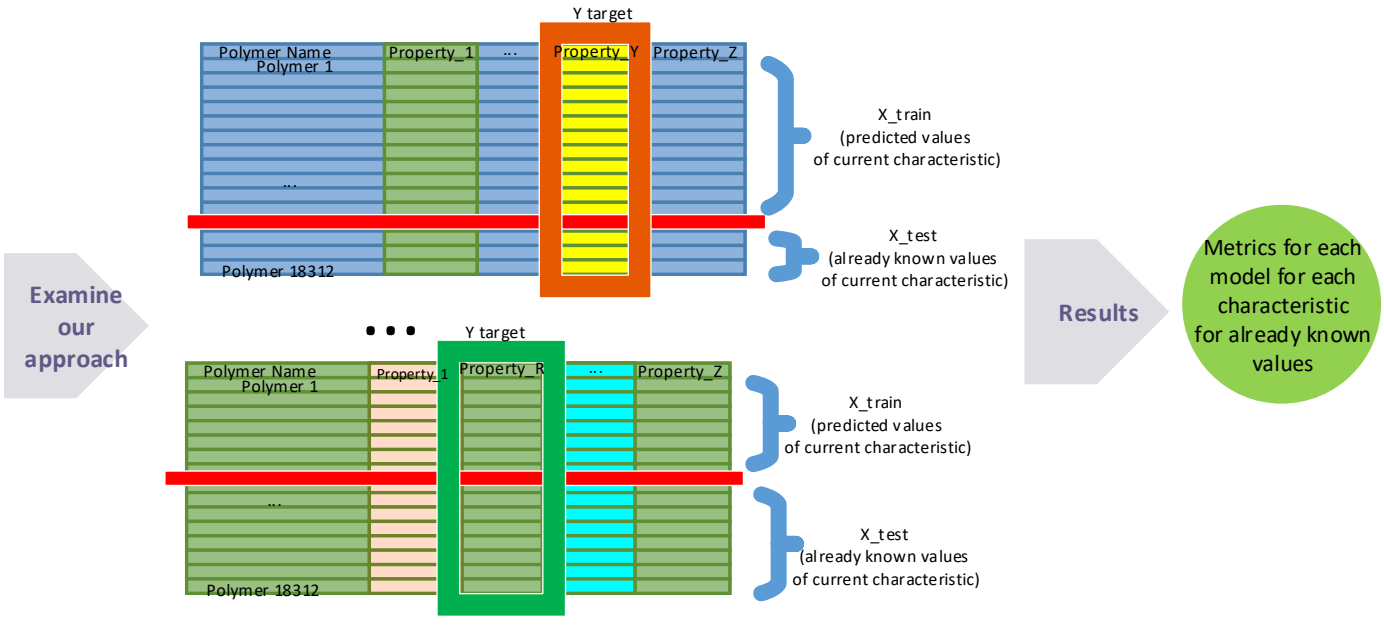

Polymer 1
Characteristics
Values
Property_1
...
...
Property_N
Polymer 2
Property_1'
...
...
Polymer K
Property_1''
...
Data transformation
Polymer Name
SMILES
Property_1
...
Property_98
Polymer 1
...
Polymer 18312


Data Prepare for ML fitting
Polymer Name
1024 binary cols
Property_A
Polymer G
Polymer Q
Polymer Name
1024 binary cols
Property_F
Polymer U
Polymer W
11 models, 68 characteristics (if non-null counts > 50)


ML fitting
80% train
20% test
params
Results
Metrics for each model for each characteristic
&
Models are being saved by Joblib


Dataset with filled values for each characteristic obtained with regression model which had been chosen by the max R2 score
Chosing Best models
Polymer Name
Polymer 1
SMILES
Property_1
...
Property_98
...
Polymer 18312


Examine our approach
Y target
Polymer Name
Polymer 1
Property_1
...
Property_Y
Property_Z
Polymer 18312
X_train
(predicted values
of current characteristic)
X_test
(already known values
of current characteristic)
Y target
Polymer Name
Polymer 1
Property_1
Property_R
...
Property_Z
Polymer 18312
X_train
(predicted values
of current characteristic)
X_test
(already known values
of current characteristic)
Results
Metrics for each model for each characteristic for already known values