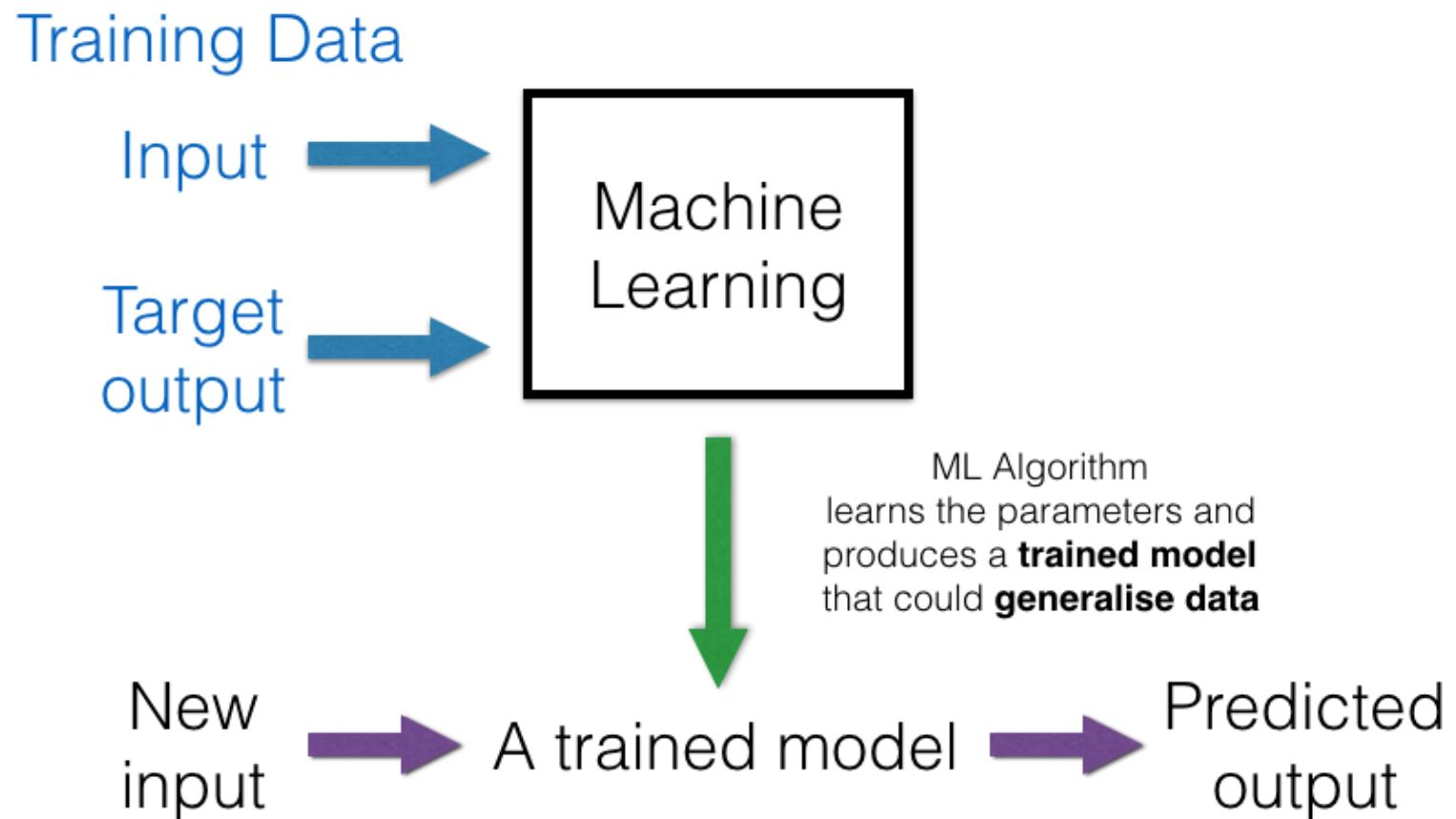




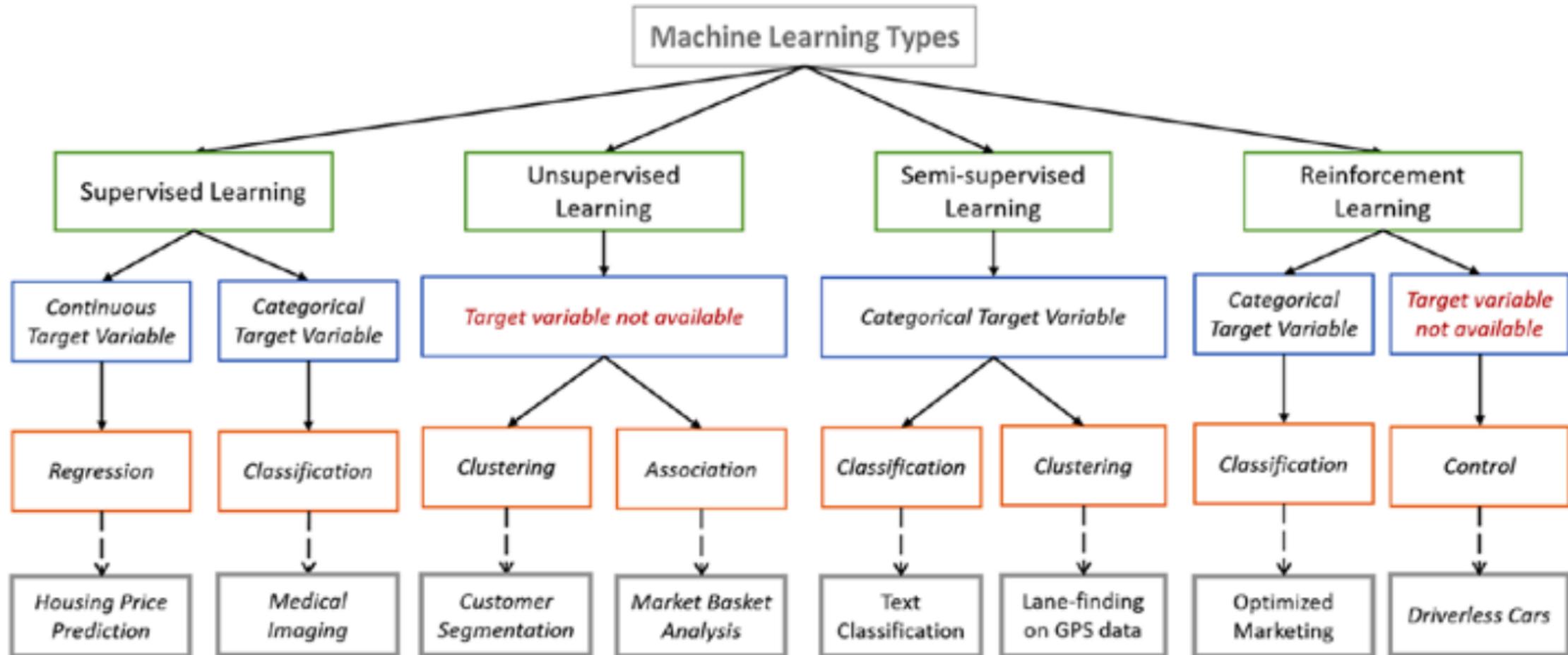
CLASSIFY DATA USING THE CLASSIFICATION LEARNER APP

PRESENTING BY POUYA MEHRBABAK

TRAINING DATA

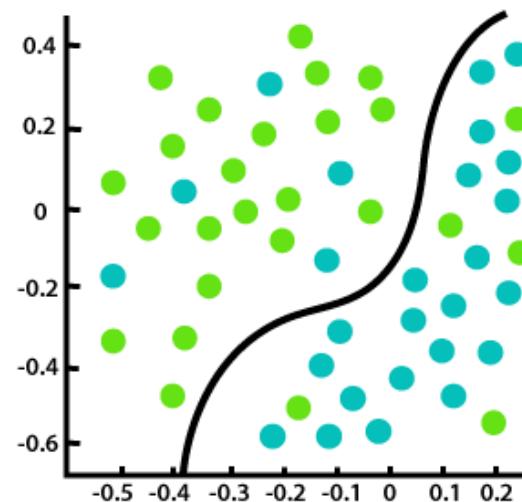


MACHINE LEARNING TYPES

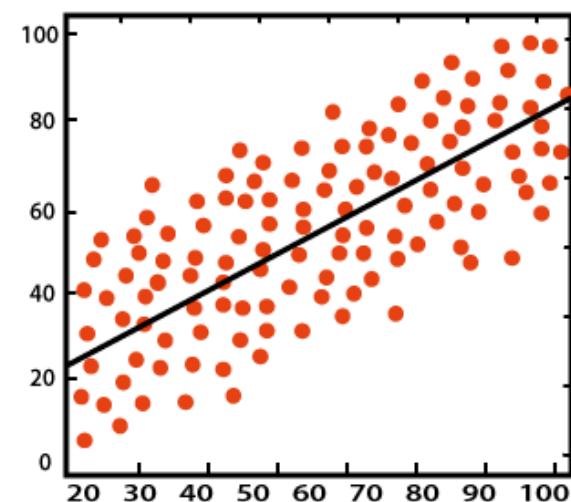


WHAT IS CLASSIFICATION IN MACHINE LEARNING ?

- **Classification**
- is a process of categorizing a given set of data into classes
- It can be performed on both structured or unstructured data.
- The process starts with predicting the class of given data points.
- The classes are often referred to as target, label or categories.

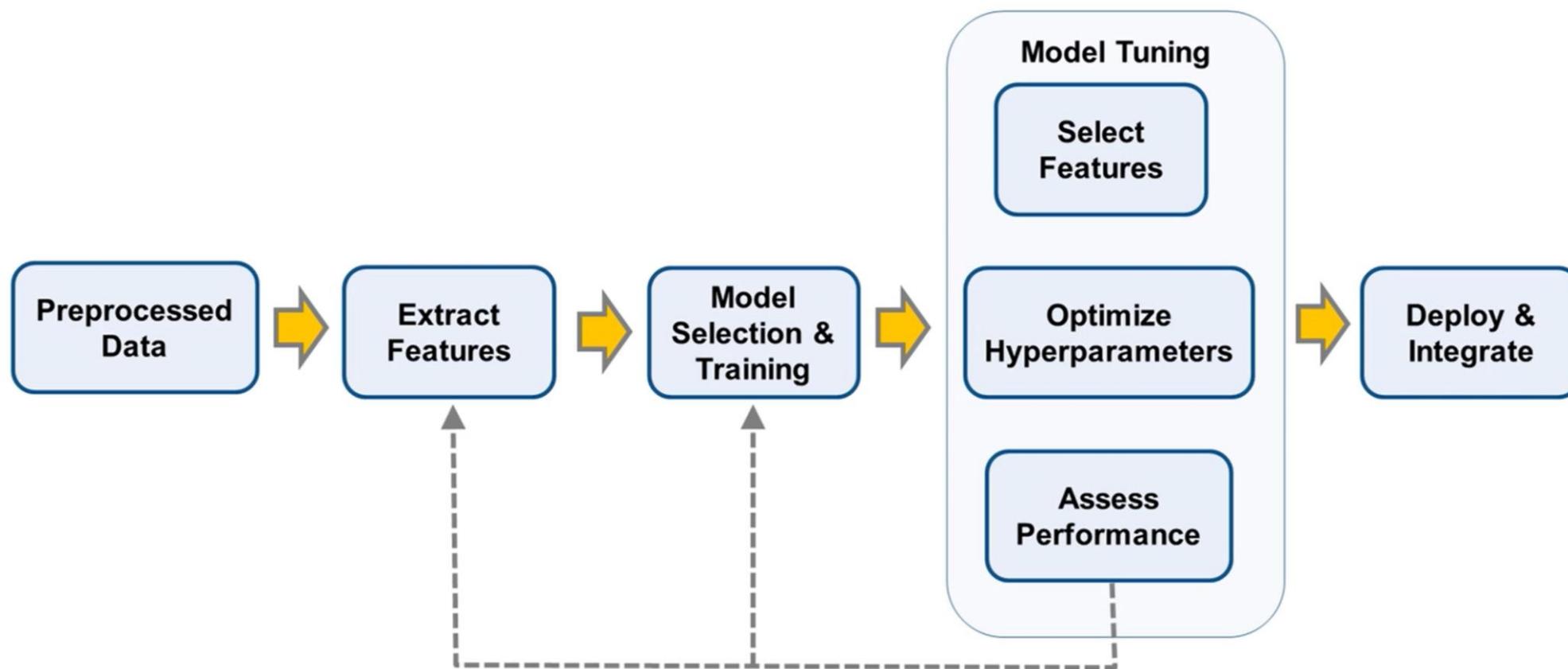


Classification



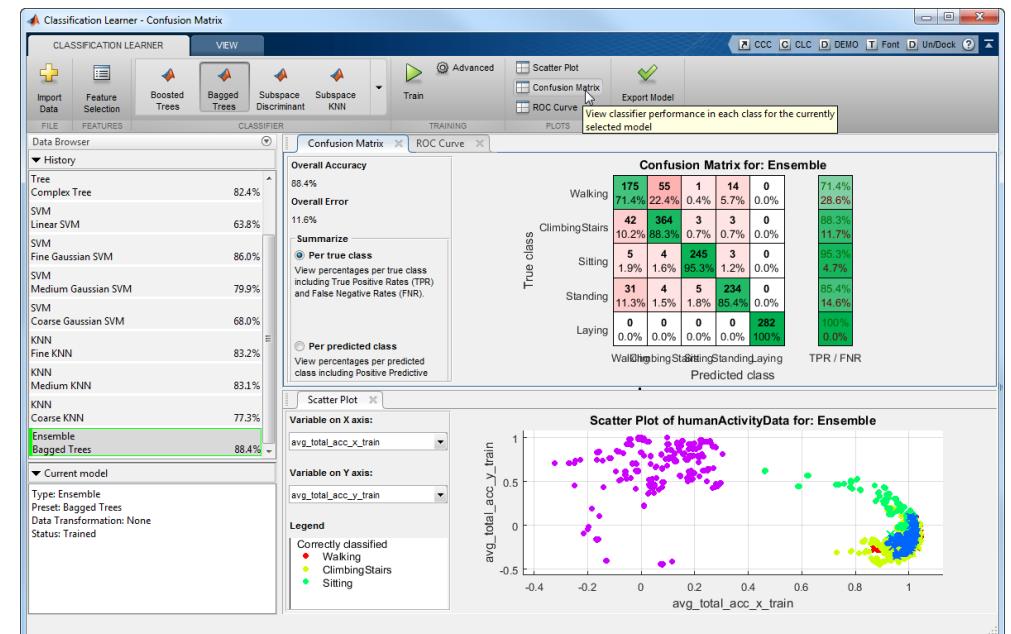
Regression

THE WORK FLOW

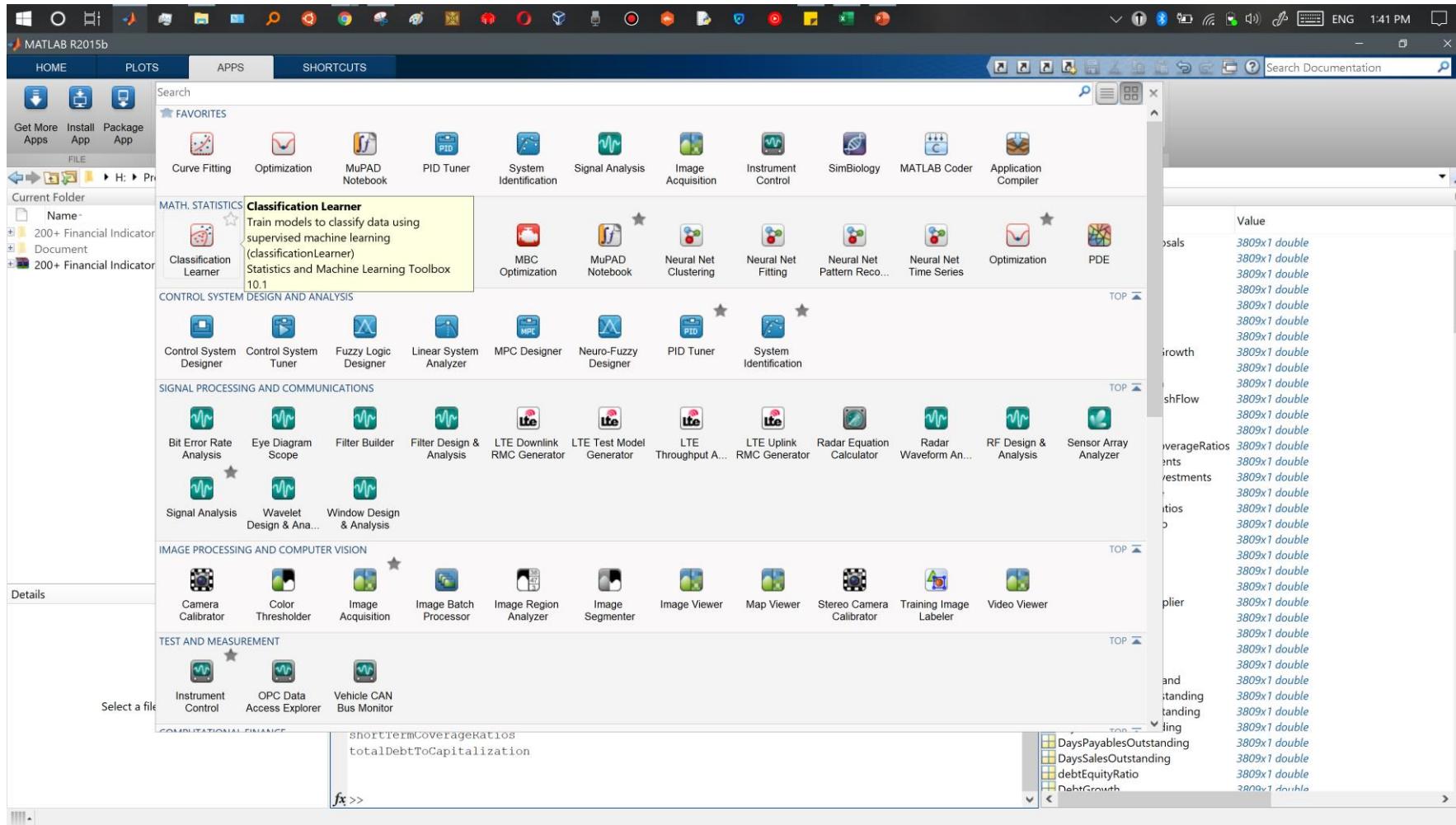


CLASSIFICATION LEARNER APP

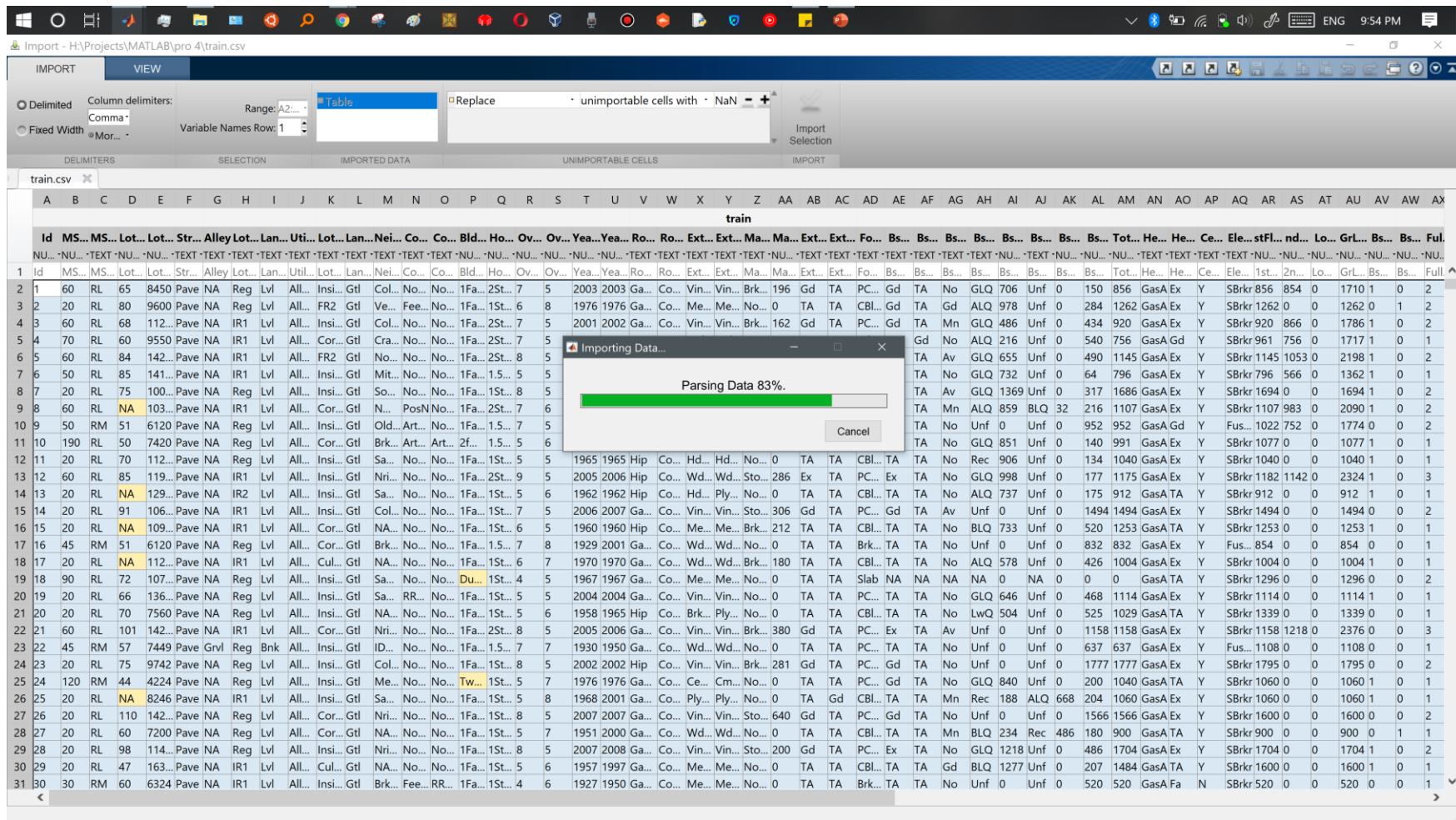
- Classification Learner is a new app in the statistics and machine learning tool box that lets you train models to classify data using supervised machine learning.
- Classification Learner lets you perform common machine learning tasks, such as:
 - importing data
 - specifying validation schemes
 - interactively exploring your data
 - selecting features
 - training models
 - assessing model performance



CLASSIFICATION LEARNER APP



SELECTING AND IMPORTING DATA



PREDICTORS AND RESPONSES

The screenshot shows the MATLAB interface with the Classification Learner app open. The title bar indicates it's a 'New Session'. The main window has three tabs: Step 1, Step 2, and Step 3.

Step 1: Select dataset from MATLAB workspace. The dataset 'train' is selected. Below it, there are two radio button options: 'Use columns as variables' (selected) and 'Use rows as variables'.

Step 2: Select predictors and response. A table lists various features with their types and ranges. The 'MSSZoning' column is highlighted in blue. The table includes columns for Name, Type, Range, and Import as.

Name	Type	Range	Import as
Id	double	1 .. 1460	Predictor
MSSubClass	double	20 .. 190	Predictor
MSSZoning	cell	5 unique	Response
LotFrontage	double	21 .. 313	Predictor
LotArea	double	1300 .. 215245	Predictor
Street	cell	2 unique	Predictor
Alley	cell	3 unique	Predictor
LotShape	cell	4 unique	Predictor
LandContour	cell	4 unique	Predictor
Utilities	cell	2 unique	Predictor
LotConfig	cell	5 unique	Predictor
LandSlope	cell	3 unique	Predictor
Neighborhood	cell	25 unique	Predictor
Condition1	cell	9 unique	Predictor
Condition2	cell	8 unique	Predictor
BldgType	double	1 .. 2	Predictor
HouseStyle	double	1 .. 2.5	Predictor
OverallQual	double	1 .. 10	Predictor
OverallCond	double	1 .. 9	Predictor
YearBuilt	double	1872 .. 2010	Predictor
YearRemodAdd	double	1950 .. 2010	Predictor
RoofStyle	cell	6 unique	Predictor
RoofMatl	cell	8 unique	Predictor
Exterior1st	cell	15 unique	Predictor
Exterior2nd	cell	16 unique	Predictor
MasVnrType	cell	5 unique	Predictor
MasVnrArea	double	0 .. 1600	Predictor
ExterQual	cell	4 unique	Predictor
ExterCond	cell	5 unique	Predictor
Foundation	cell	6 unique	Predictor

Step 3: Define validation method. The 'Cross Validation' option is selected. It states: 'Protects against overfitting by partitioning the data set into folds and estimating accuracy on each fold.' Below this, it says 'Cross-validation folds: 5 folds' with a dropdown menu showing '1' through '5'. There are also options for 'Holdout Validation' (with a 'Percent held out: 25%' field) and 'No Validation'.

At the bottom of the dialog are 'Start Session' and 'Cancel' buttons.

At the very bottom of the screen, status bars show: 'Original Dataset: Observations: 0 Predictors: 0 Response Variable: Response Classes: 0 Size of Dataset: 0 B' and 'Validation: No Validation'.

PREDICTORS AND RESPONSES

- **Predictors** are the features that we use to classify them into the respective classes.
- The **response** indicated here is the parameter into which we will classify the data.

Step 2

Select predictors and response.

Name	Type	Range	Import as
Id	double	1 .. 1460	Predictor
MSSubClass	double	20 .. 190	Predictor
MSZoning	cell	5 unique	Response
LotFrontage	double	21 .. 313	Predictor
LotArea	double	1300 .. 215245	Predictor
Street	cell	2 unique	Predictor
Alley	cell	3 unique	Predictor
LotShape	cell	4 unique	Predictor
LandContour	cell	4 unique	Predictor
Utilities	cell	2 unique	Predictor
LotConfig	cell	5 unique	Predictor
LandSlope	cell	3 unique	Predictor
Neighborhood	cell	25 unique	Predictor
Condition1	cell	9 unique	Predictor
Condition2	cell	8 unique	Predictor
BldgType	double	1 .. 2	Predictor
HouseStyle	double	1 .. 2.5	Predictor
OverallQual	double	1 .. 10	Predictor
OverallCond	double	1 .. 9	Predictor
YearBuilt	double	1872 .. 2010	Predictor
YearRemodAdd	double	1950 .. 2010	Predictor
RoofStyle	cell	6 unique	Predictor
RoofMatl	cell	8 unique	Predictor
Exterior1st	cell	15 unique	Predictor
Exterior2nd	cell	16 unique	Predictor
MasVnrType	cell	5 unique	Predictor
MasVnrArea	double	0 .. 1600	Predictor
ExterQual	cell	4 unique	Predictor
ExterCond	cell	5 unique	Predictor
Foundation	cell	6 unique	Predictor

VALIDATION

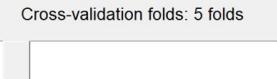
- **Validation:** the training data is divided into training and validation parts.
 - The validation part is used to check the trained model before testing to see that the model doesn't get overfitted.
 - **Cross-Validation**
 - In this, the number of folds is the number of parts the training data will randomly divide. One part of it will be used for validation and the remaining for training.
 - **Holdout Validation**
 - We can select the percentage of data held out as a testing dataset and the remaining for training. This is usually recommended for large datasets.

Step 3

Define validation method.

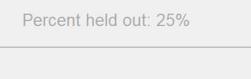
Cross Validation

Protects against overfitting by partitioning the data set into folds and estimating accuracy on each fold.



Holdout Validation

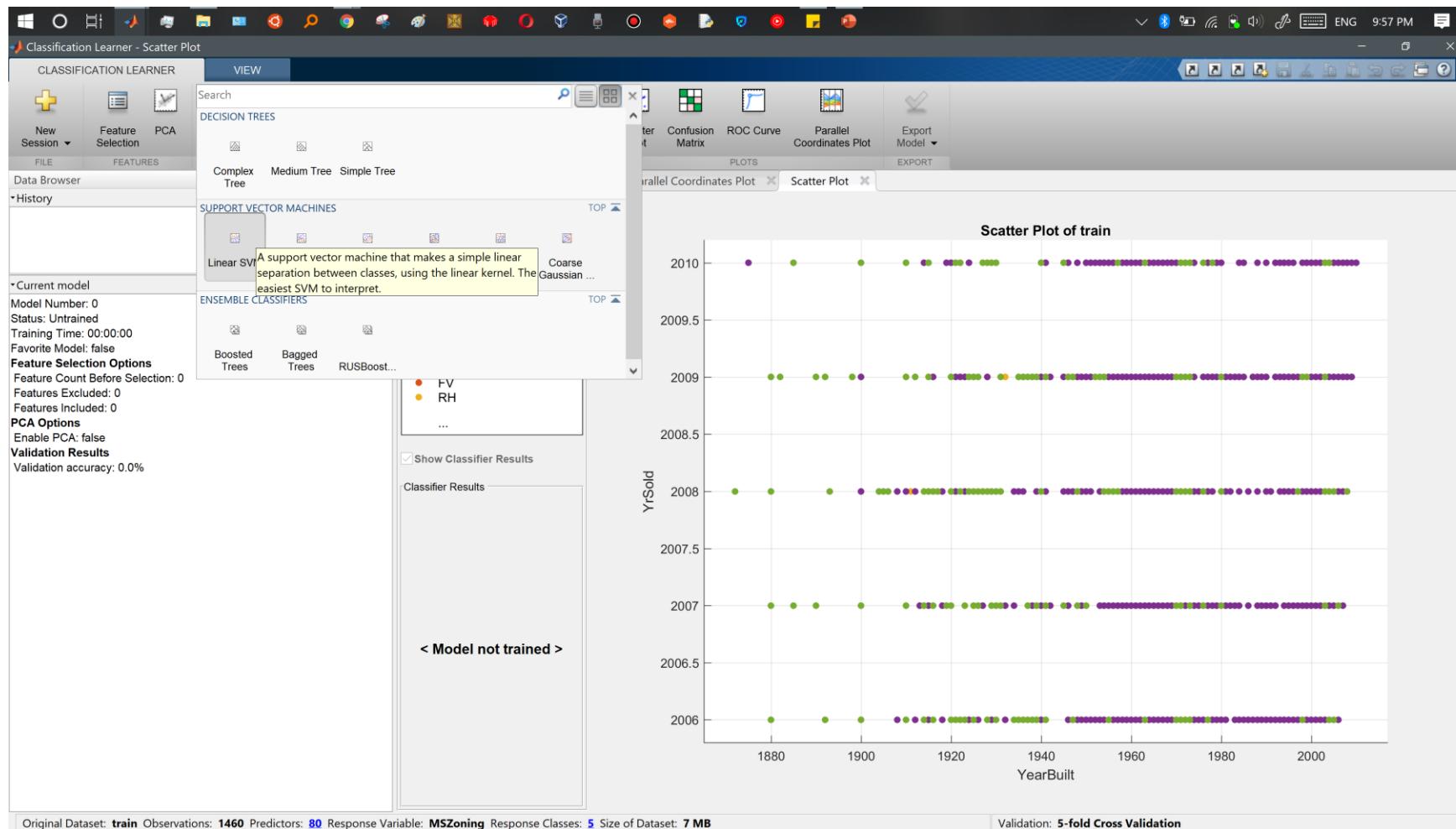
Recommended for large data sets.



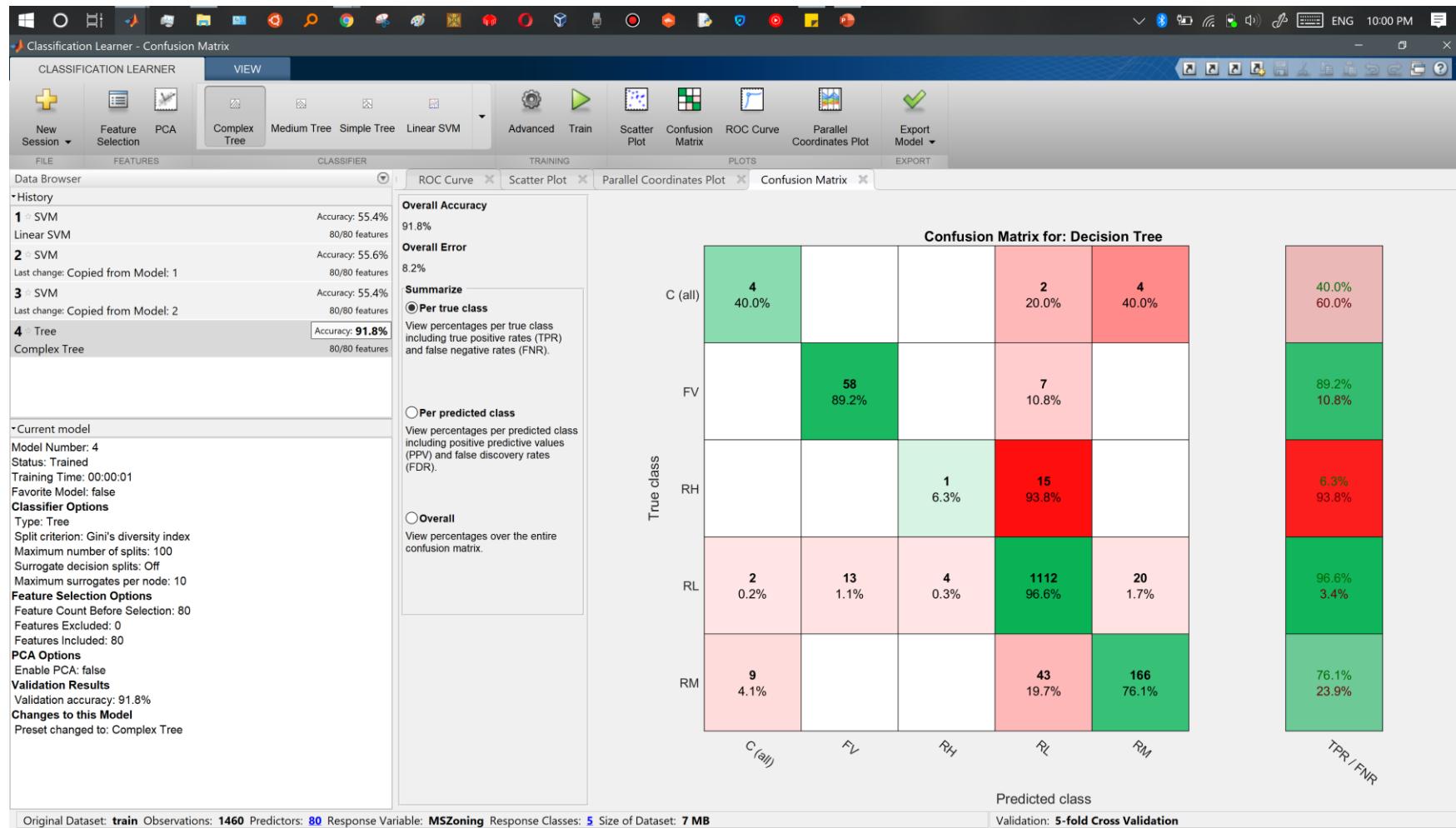
No Validation

No protection against overfitting.

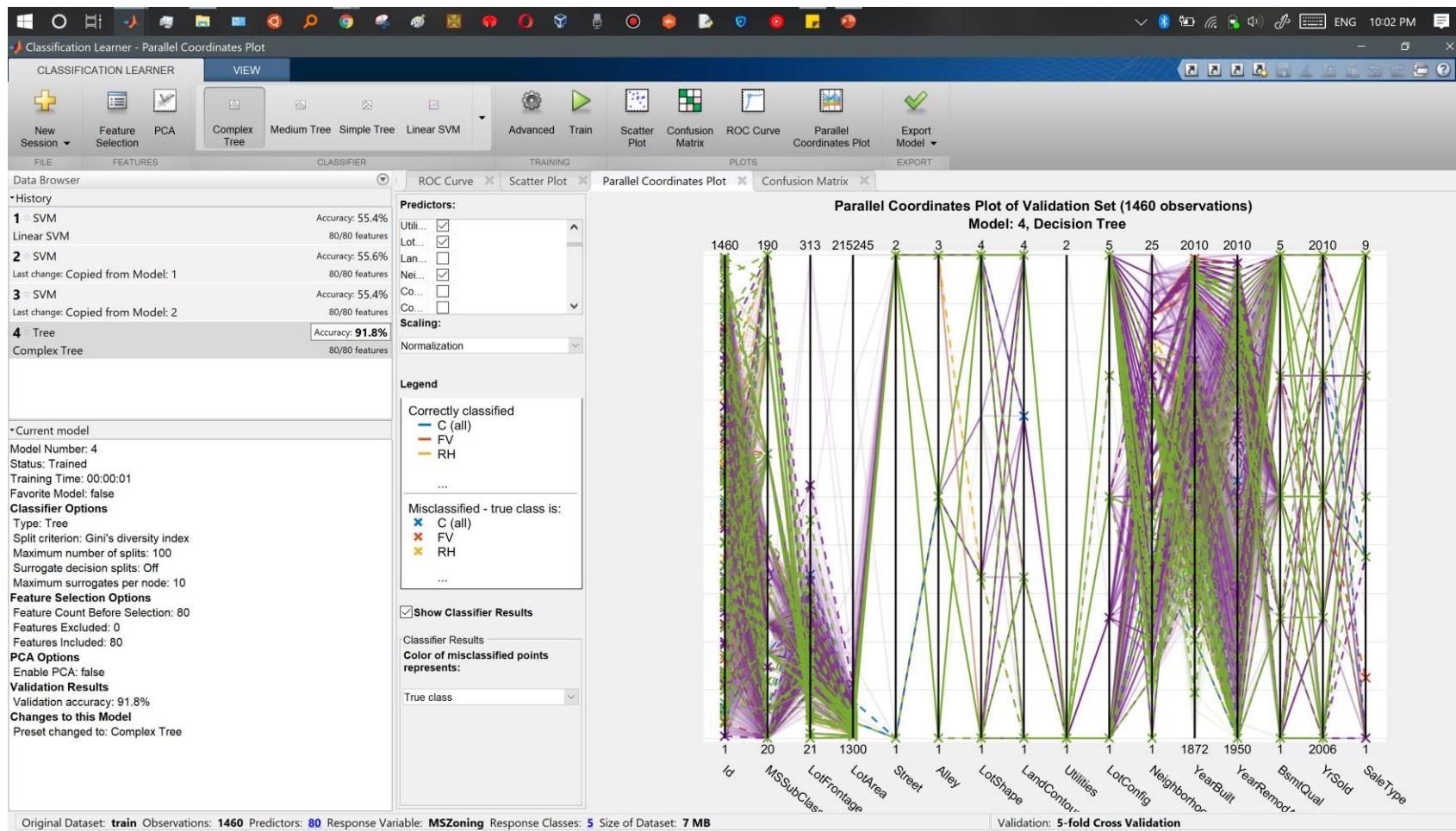
CLASSIFICATION LEARNER APP



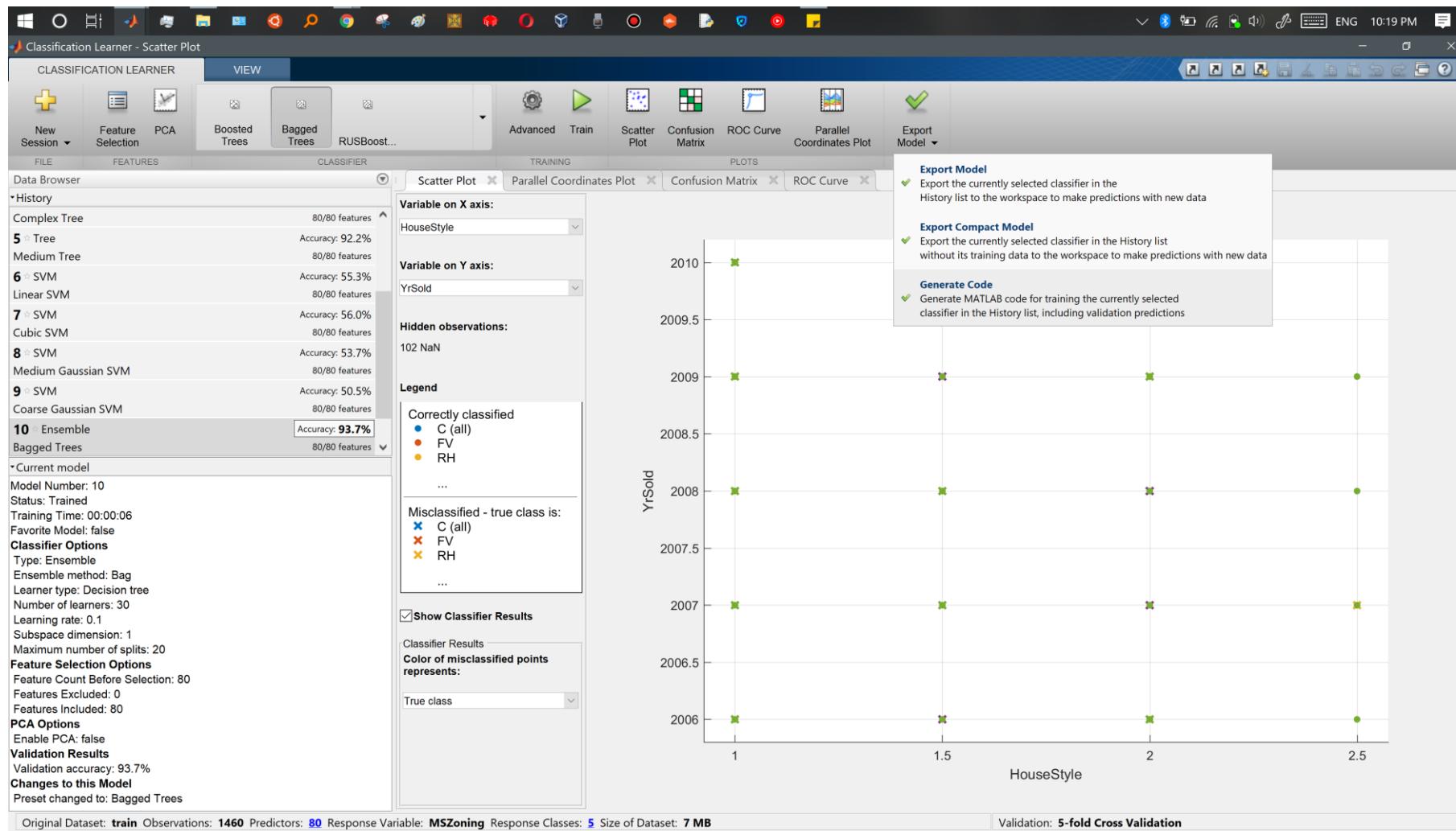
CONFUSION MATRIX



PARALLEL COORDINATES PLOT OF VALIDATION SET



EXPORT MODEL AND GENERATE CODE



SOURCES AND LINKS

- Source :
 - <https://www.mathworks.com/help/stats/classificationlearner-app.html>
 - <https://matlabhelper.com/blog/matlab/classification-learner-app-in-matlab/>
 - <https://www.youtube.com/watch?v=3YGtxY67bq0>
 - <https://www.kaggle.com/anniepyim/essential-classification-algorithms-explained>



THANK YOU

POUYA MEHRBABAK