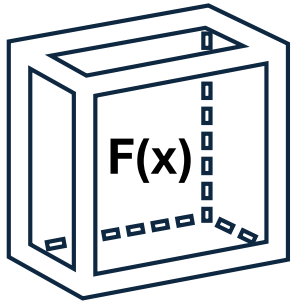
Machine Learning

Lecture Hub

By Promise Ekeh

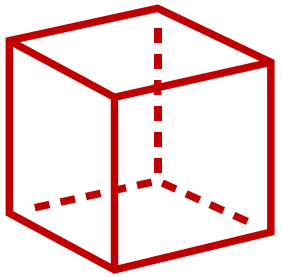
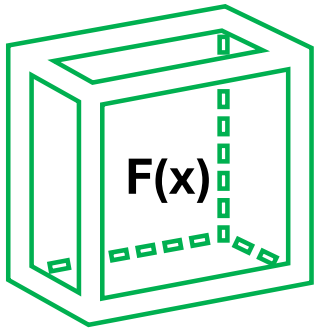
Machine Learning

Traditional programming

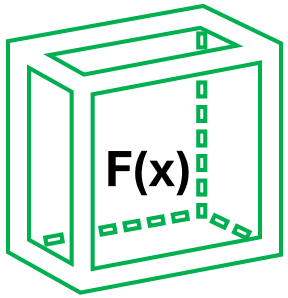
[1, 2, 3] → 

→ [2, 4, 6]

Machine Learning

[1, 2, 3]
[2, 4, 6] →  → 

Model training
model.fit()

[5, 6, 7] → 

→ [10, 12, 14]

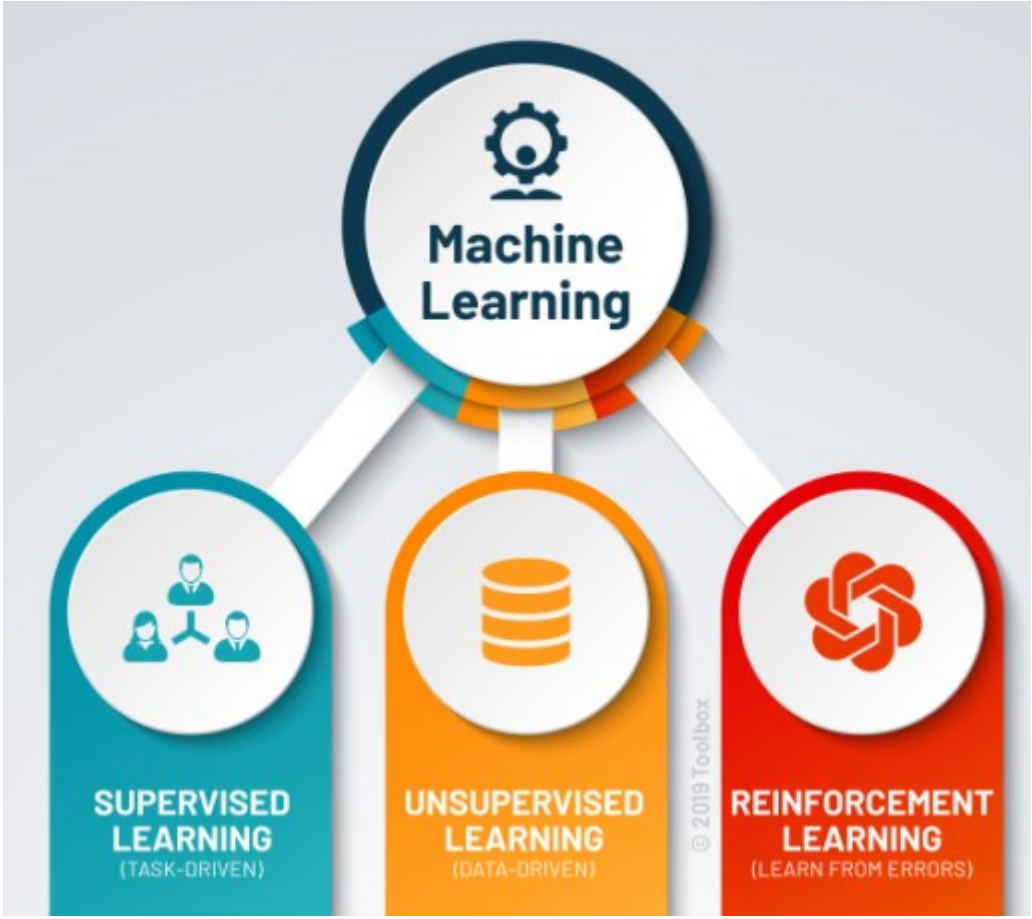
Making prediction
model.predict()

Types of Machine Learning

Supervised learning: ML algorithm is trained on labelled data. The training dataset is also very similar to the final dataset in its characteristics.

Unsupervised learning involves discovering patterns and structures in unlabelled data. Unlike supervised learning, there is no predefined target variable.

Reinforcement learning directly takes inspiration from how human beings learn from data in their lives. It features an algorithm that improves upon itself and learns from new situations using a trial-and-error method. Favourable outputs are encouraged or '*reinforced*', and non-favourable outputs are discouraged or '*punished*'.



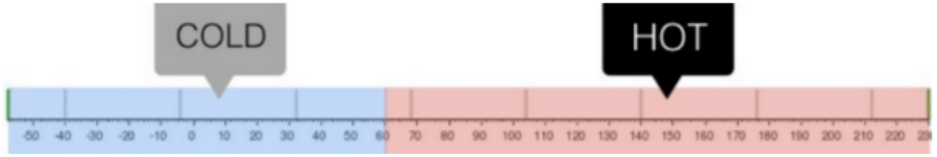
Supervised Machine Learning

Supervised learning divided into two categories:

Classification algorithms are used when the output variable is categorical, such as classifying emails as spam or non-spam, identifying handwritten digits, or predicting the sentiment of customer reviews. The model learns from labelled examples to classify new data into predefined categories.

Regression algorithms are employed when the output variable is continuous or numerical, like predicting house prices based on various features or estimating the sales of a product. The model learns from labelled examples to make predictions within a range of values.

Will it be hot or cold tomorrow?

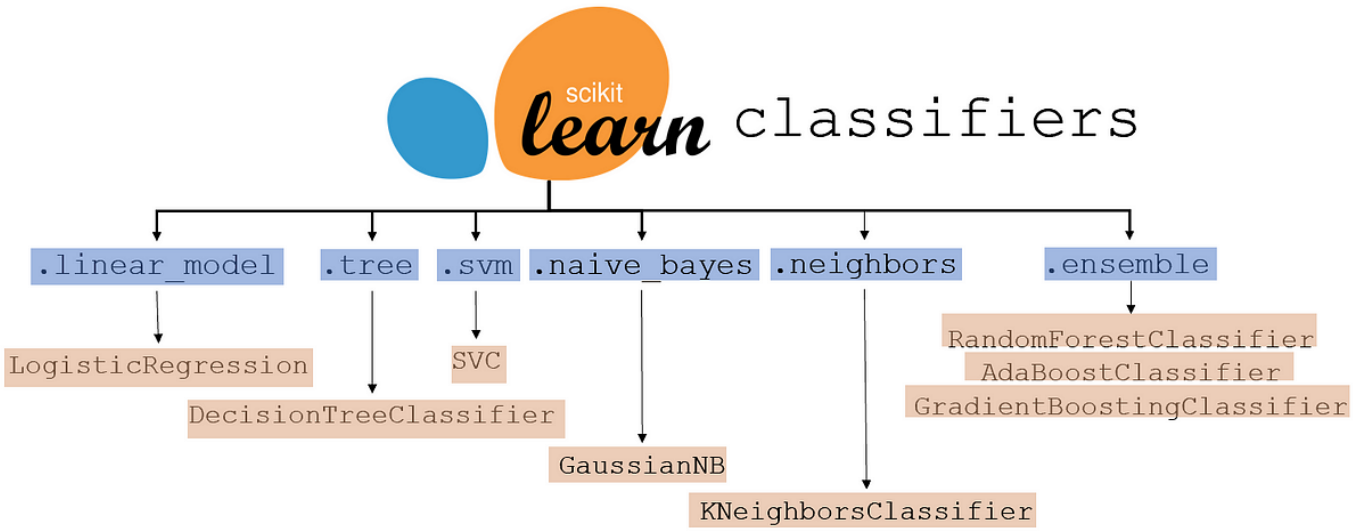


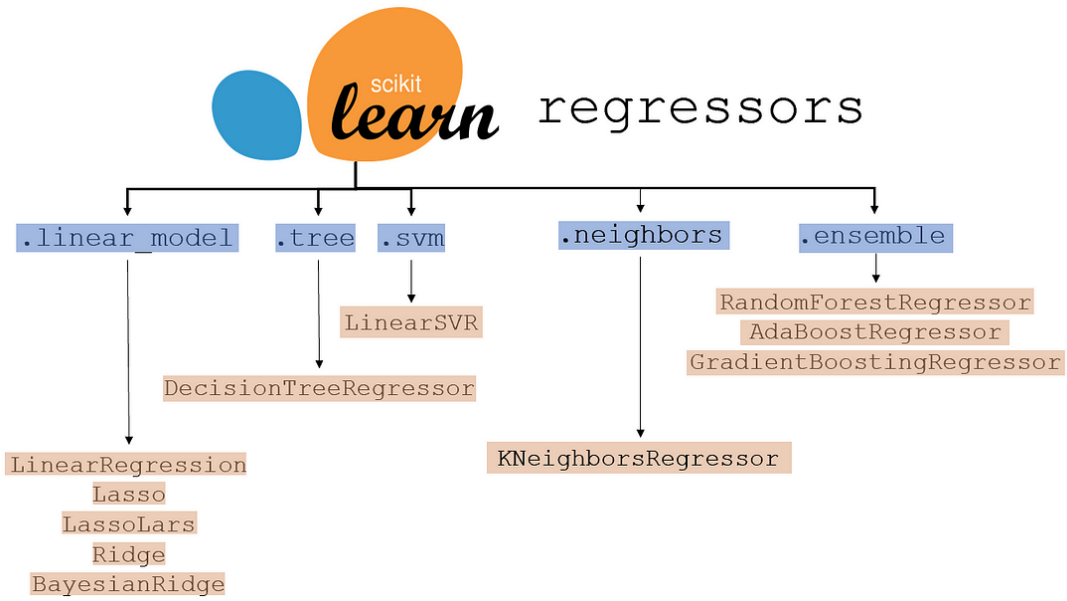
Fahrenheit

What will be the temperature tomorrow?

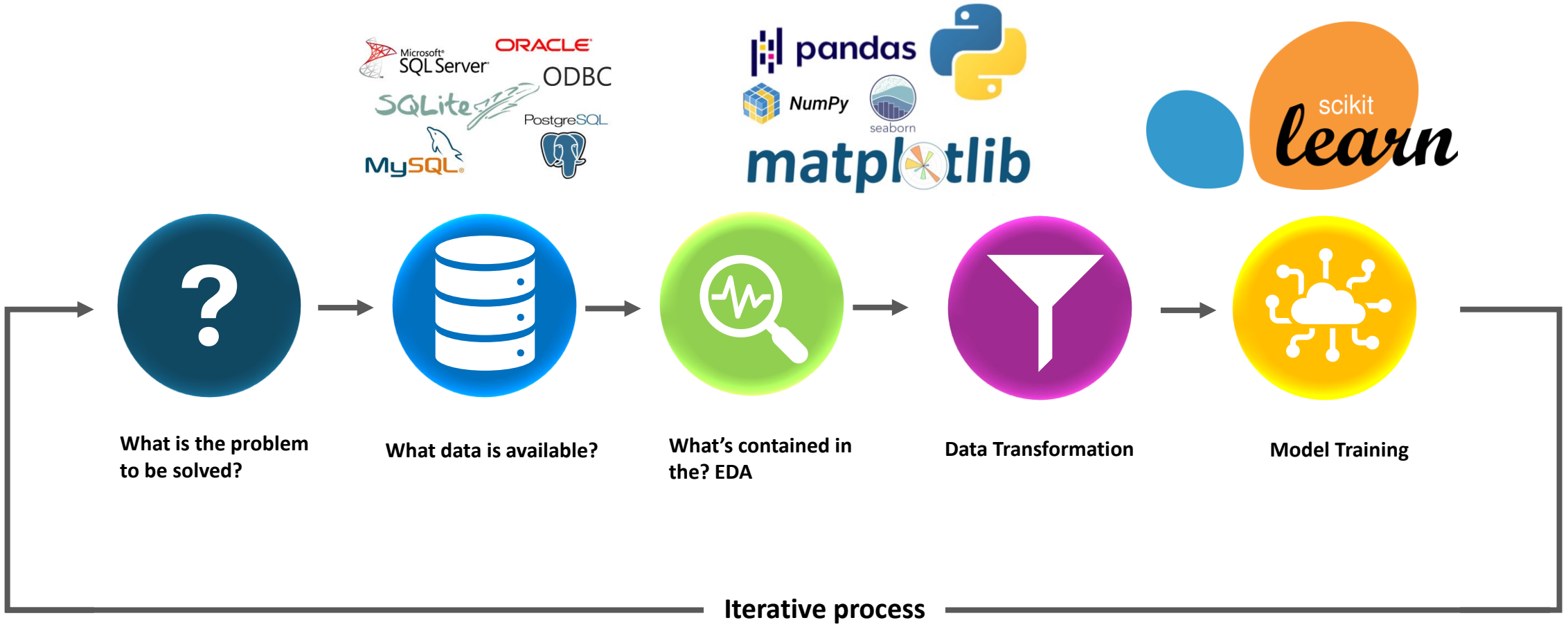
Fahrenheit

Supervised Machine Learning Models

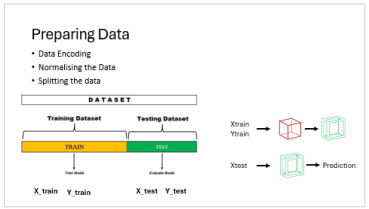


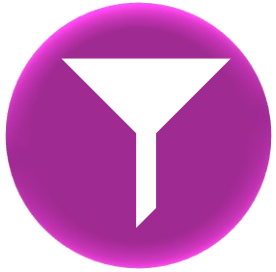


Workflow


Microsoft SQL Server
ORACLE
ODBC
SQLite
PostgreSQL
MySQL
pandas
NumPy
seaborn
matplotlib
scikit learn
What is the problem to be solved?
What data is available?
What's contained in the? EDA
Data Transformation
Model Training
Iterative process

Model Training Workflow





Data Transformation

Model Training

Choosing the model

Fitting the model

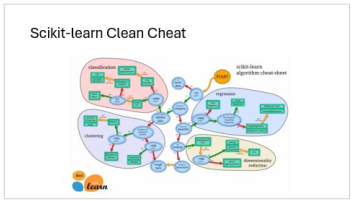
Making Prediction

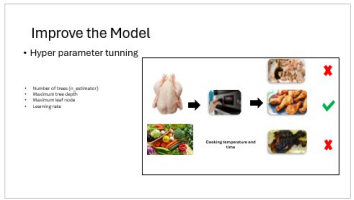
Evaluate model Prediction

Improve model Prediction

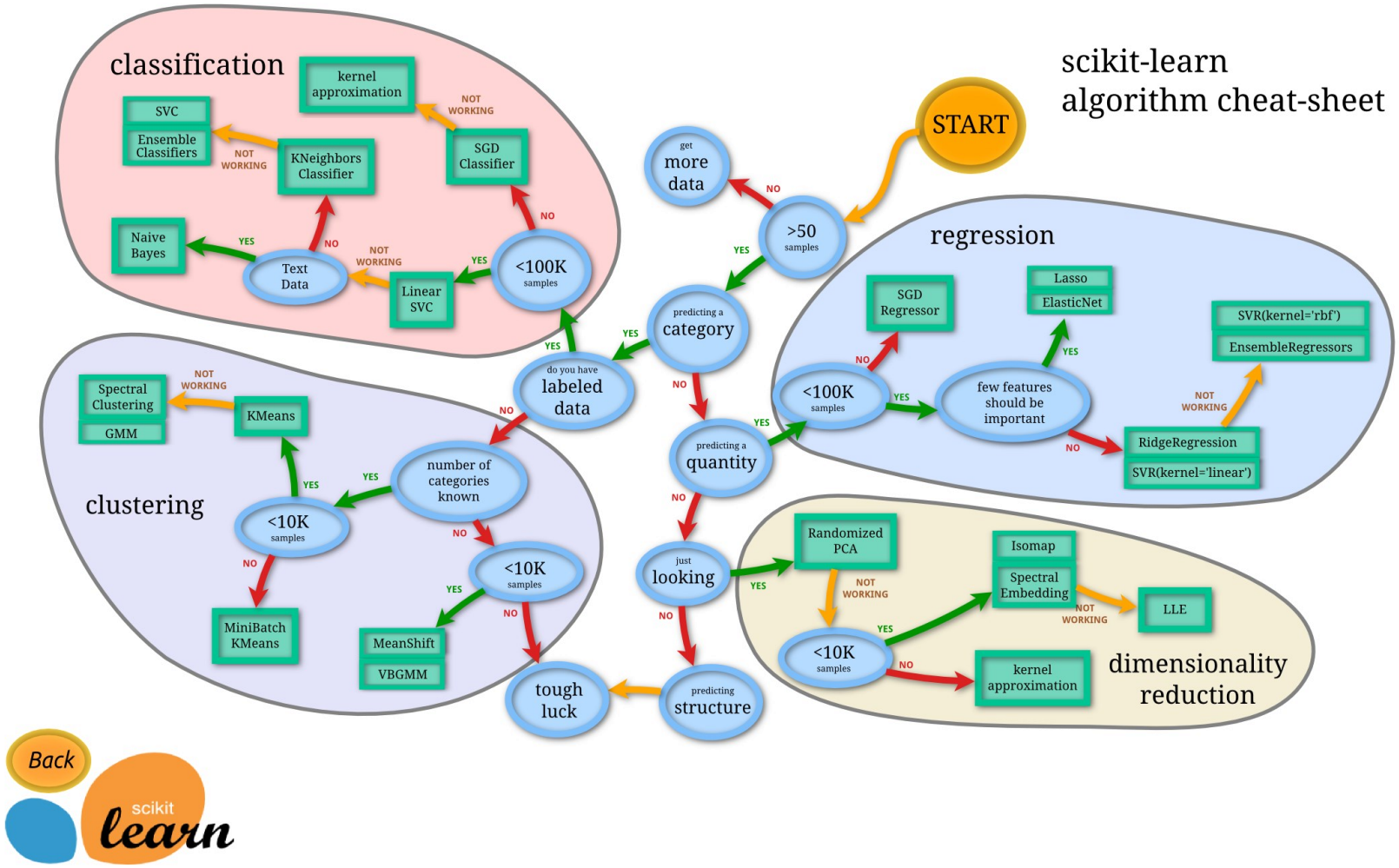
Save & Load Model

Deploy Model



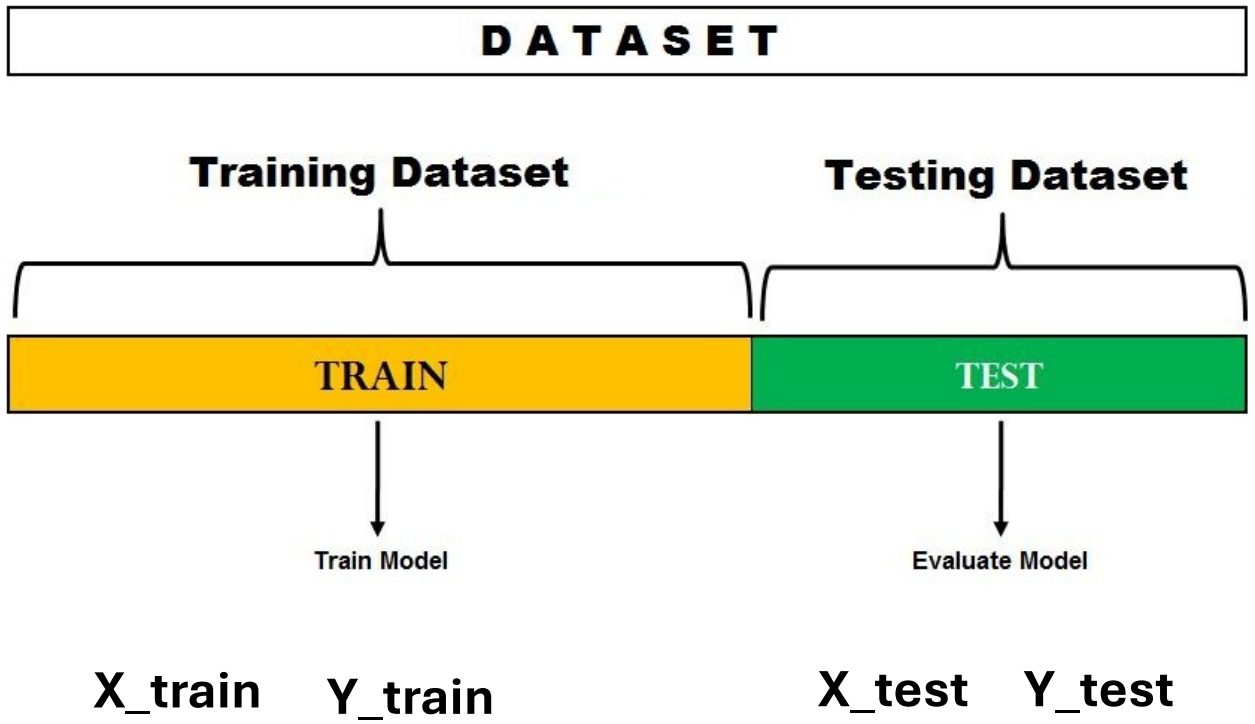


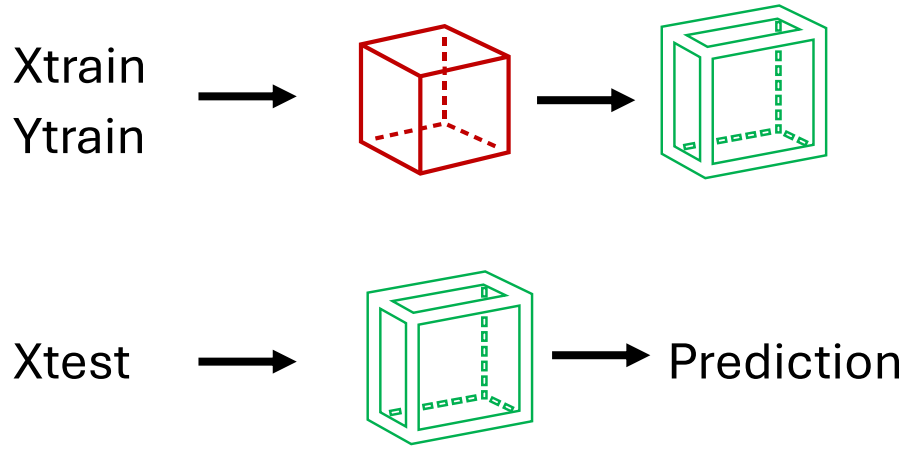
Scikit-learn Clean Cheat



Preparing Data

- Data Encoding
- Normalising the Data
- Splitting the data





Model Evaluation

commonly used metrics marked in green

metrics

classification

- accuracy_score
- balanced_accuracy_score
- average_precision_score
- brier_score_loss
- f1_score
- log_loss
- precision_score
- recall_score
- jaccard_score
- roc_auc_score

regression

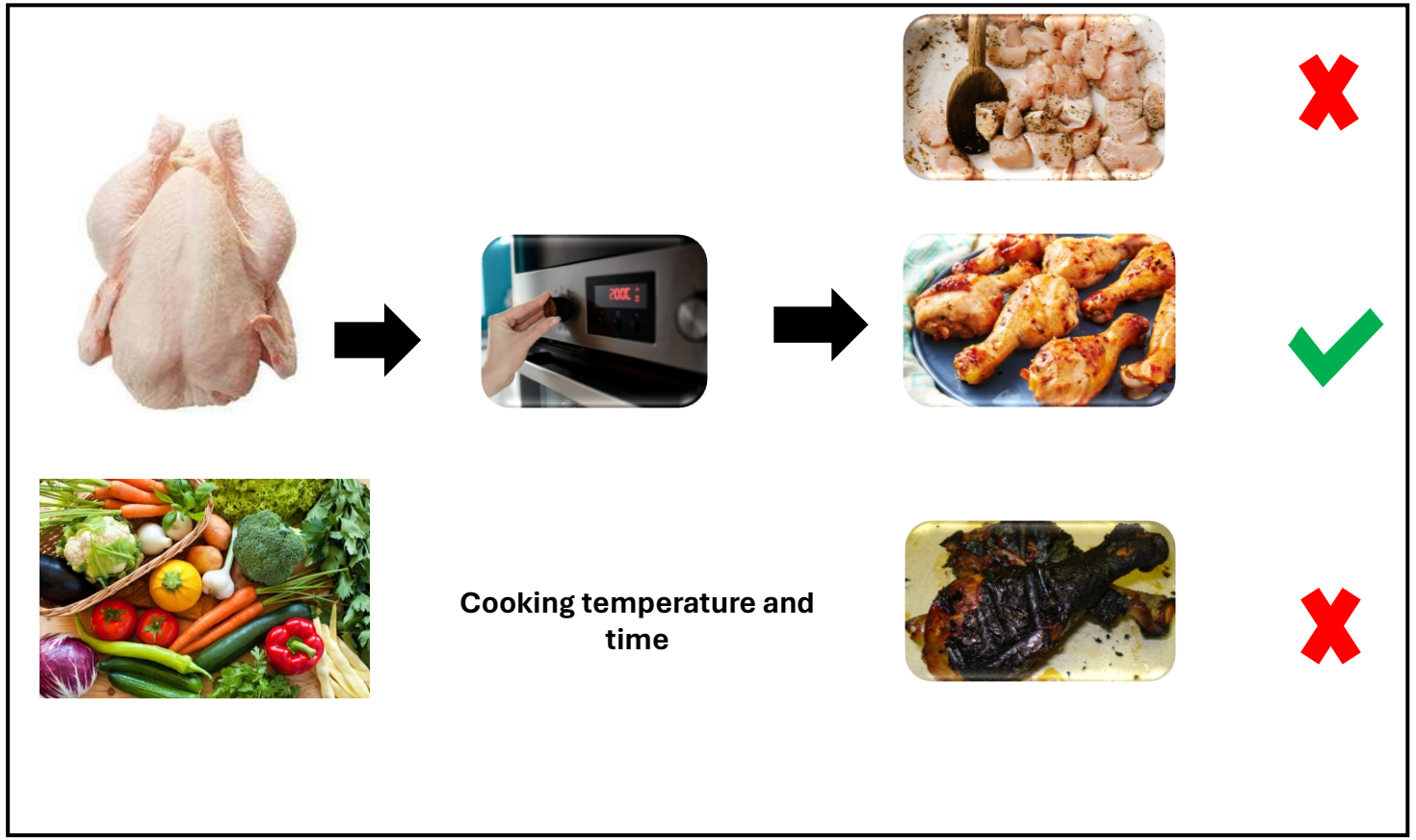
- explained_variance_score
- max_error
- mean_absolute_error
- mean_squared_error
- mean_squared_log_error
- median_absolute_error
- r2_score
- mean_poisson_deviance
- mean_gamma_deviance

Prediction Y_test

Improve the Model

- Hyper parameter tunning

  - Number of trees (n_estimator)
  - Maximum tree depth
  - Maximum leaf node
  - Learning rate



Over fitting vs under fitting

Overfitting

When your model learns all complex and noise from training data and performs well in training data but while coming to validation data it does not work well then our data is overfitting.

Good performance on the training data, poor generliazation to other data.

Underfitting

When our data is underfitting then our model does learn the underlying trend data. It occurs when we have fewer data to build the model or when we try to build the linear model with non-linear data. :

Poor performance on the training data and poor generalization to other data