Below is the workflow of a ml project derived from 《Hands On Machine Learning with Scikit Learn and TensorFlow》

1. Look at the big picture.

Frame the problem

Select a Performance Measure

Check the Assumption

Create a isolated env

2. Get the data.

3. Discover and visualize the data to gain insights. （EDA）

Visualizing

Correlation

Attribute combinations

4. Prepare the data for Machine Learning algorithms. （Feature Engineering）

"""

You should write functions to do that, for several good reasons:

• Reproduce these transformations easily on any dataset (e.g., the next time you get a fresh dataset).

• You will gradually build a library of transformation functions that you can reuse in future projects.

• You can use these functions in your live system to transform the new data before feeding it to your algorithms.

• This will make it possible for you to easily try various transformations and see which combination of transformations works best.

"""

Data Cleaning

Handling Text and Categorical Attributes

Custom Transformers

Feature scaling

Transformation Pipelines

5. Select a model and train it.

"""

At last! You framed the problem, you got the data and explored it, you sampled a training set and a test set,

and you wrote transformation pipelines to clean up and prepare your data for Machine Learning algorithms automatically.

You are now ready to select and train a Machine Learning model.

"""

Training and Evaluating on the Training Set

Cross Validation

6. Fine-tune your model.（Fine tune hyper parameter）

Grid Search

randomized Search

ensemble methods

analyze the best models and results

test on test set

7. Present your solution.

8. Launch, monitor, and maintain your system.