### Machine Learning Process**Preparing the Data:**

1. **Clean** the data to remove unwanted data, missing values, rows, and columns, duplicate values, data type conversion, etc.
2. [**Visualize** the data](https://www.simplilearn.com/data-visualization-article) to understand how it is structured and understand the relationship between various variables and classes present.
3. (If scaling or encoding is necessary, you can scale or encode your data at this stage)
4. **Split** the cleaned data into two sets - a training set and a testing set.

**Modeling (Training) + Evaluation:**

1. Choose your problem (Classification or Regression)
2. Try any model that fits your problem (Logistic, Linear, SVM, Decision Tree, Random Forest, AdaBoost, XGBoost, etc.)
3. Train each model with default parameters.
4. **Cross validate** each model to see if there is an overfitting or an underfitting problem.
5. Apply **Grid search** to each model to get the best parameters.
6. Then **cross validate** your model again with the best parameters you got after using Grid search.
7. You can utilize feature importance here.
8. Assign each model’s scores to variables and compare them at the end. Then choose the model that gives the best scores.

**Prediction**

1. Load your test data
2. Clean your test data to fit your train data (if necessary)
3. Create the model you choose with the best parameters.
4. Get your predictions.