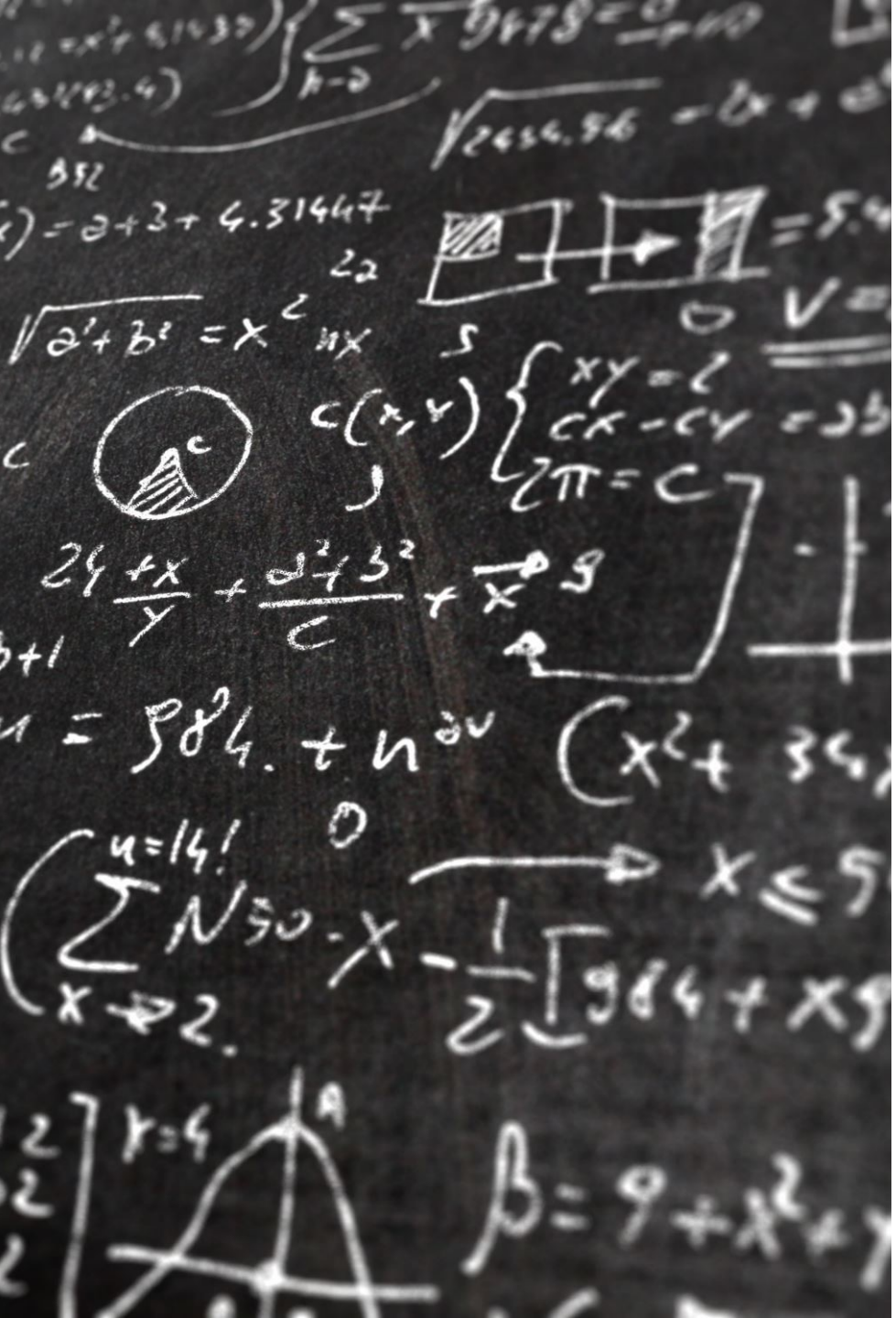
# Template for Machine Learning

# Recipe has 6 Ingredients

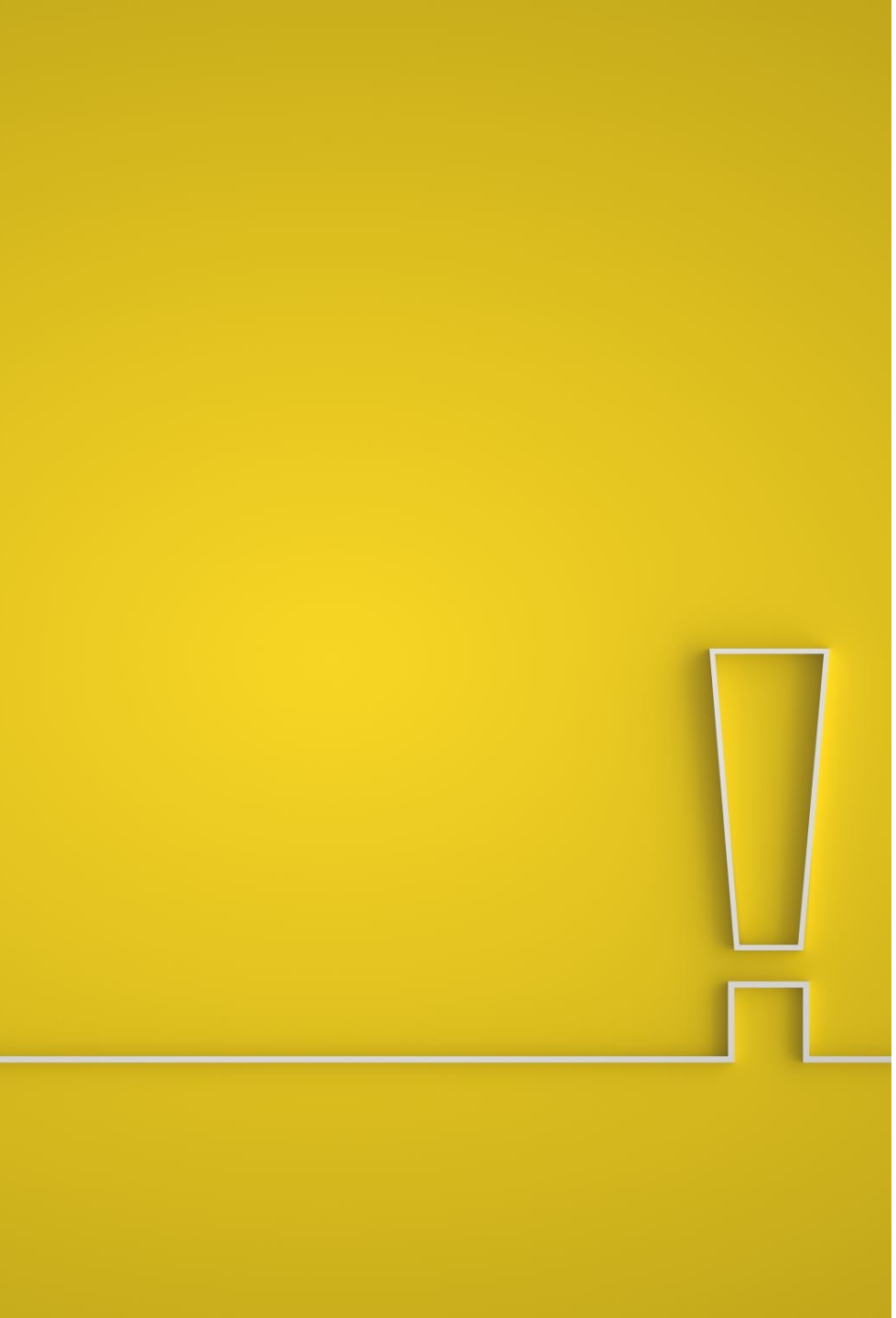




# Data

- All Data should be in numeric in nature
- Check whether we have both input (feature) and output (target)
  - If Yes → Supervised ML
    - If the output – Numeric/continuous – Regression task
    - If the Output is Categorical – Classification task
  - If No → Unsupervised ML



# Tasks

- If the Output variable is continuous – Regression
- If the output variable is categorical – classification/discrete (if I have less that 20 discrete values)
- Unsupervised :
  - Clustering – if the task is to create groups
  - Recommendation system – if the task is to recommend the items

# Model

## Classification :

- DecisionTreeClassfier
- RandomForestClassifier
- LogisticRegression
- SVC
- AdaBoostClassifier
- GradientBoostClassifier

## Regression:

- LinearRegression
- DecisionTreeRegressor
- SVR
- RandomForestRegressor
- AdaBoostRegressor
- GradientBoostRegressor

## Clustering : KNN

## Recommendation : Apriori

# Loss Function/Evaluation Function (applicable for Supervised Learning)

## Regression :

- Mean Squared Error
- Mean Absolute Error
- Root Mean Squared Error
- Mean squared log error
- R2-score

## Classification:

- Accuracy
- Entropy
- F1 Score
- Precision
- Recall
- Confusion Matrix
- Classification Report
- Gini
- ROC AUC Curve

# Steps we Follow in ML

- All the data in numerical format
- Drop the un-necessary columns
- Taking care of Missing values
  - If the number of missing rows is less than 20% - imputation
  - If missing is greater than 20% - remove the column
- Create Heatmap to understand the relationship between features and target
- Separate the Input & Output
- Split the dataset
- Perform the fit
- Generate prediction on test data
- Evaluate the model on test data

# Learning Algorithm

This Describes how a model learns the pattern in the dataset