**Stacking Discussion**

**1. Voting Classifier (voting\_clf1)**

python

CopyEdit

voting\_clf1 = VotingClassifier(

estimators=[('clf1', rf), ('clf2', dt), ('clf3', gb), ('clf4', gnb)],

voting='hard'

)

**What is a Voting Classifier?**

A **Voting Classifier** is an ensemble model that combines predictions from multiple base models and decides the final prediction based on majority voting (voting='hard') or averaging (voting='soft').

**Components in voting\_clf1**

* estimators: A list of tuples where each tuple consists of a name (string) and a classifier instance.
* voting='hard': This means the final prediction is determined by majority voting among the base classifiers.

**Base Models Used in VotingClassifier**

| **Model** | **Name in Code** | **Description** |
| --- | --- | --- |
| RandomForestClassifier | 'clf1' | Uses multiple decision trees to make predictions by averaging their outputs |
| DecisionTreeClassifier | 'clf2' | A single decision tree that makes predictions based on feature splits |
| GradientBoostingClassifier | 'clf3' | Uses boosting (iteratively improving weak models) to create a strong predictor |
| GaussianNB | 'clf4' | A probabilistic classifier based on Bayes' theorem |

**How It Works**

1. Each of the 4 classifiers (rf, dt, gb, gnb) makes a prediction on an input data point.
2. Since voting='hard', the class that gets the most votes from these models is selected as the final prediction.

**2. Stacking Classifier (sclf1)**

python

CopyEdit

sclf1 = StackingClassifier(

estimators=[('clf1', rf), ('clf2', dt), ('clf3', gb), ('clf4', gnb)],

final\_estimator=voting\_clf1

)

**What is a Stacking Classifier?**

A **Stacking Classifier** is another ensemble model that combines multiple base classifiers and trains a meta-classifier to make the final prediction.

**Components in sclf1**

* estimators: Same 4 base models as used in VotingClassifier.
* final\_estimator=voting\_clf1: The meta-classifier is the Voting Classifier (voting\_clf1).

**How Stacking Works**

1. **First Layer (Base Models)**:
   * The base models (rf, dt, gb, gnb) independently make predictions on the dataset.
   * Their predictions are used as **input features** for the next step.
2. **Second Layer (Meta-Classifier)**
   * Instead of making a final decision directly, the outputs from the base classifiers are fed into another model (voting\_clf1).
   * This meta-classifier combines their results and makes the final decision.

**Why Use Stacking?**

* Unlike voting, which just selects the most frequent prediction, **stacking learns a new model** (meta-classifier) based on the base models’ outputs.
* The meta-classifier can **correct mistakes** made by individual models by learning patterns in their predictions.

**Final Workflow of the Code**

1. The **Voting Classifier (voting\_clf1)** is trained using **4 base classifiers**.
2. The **Stacking Classifier (sclf1)** is created using the **same 4 base classifiers** and a **Voting Classifier as the meta-classifier**.
3. Predictions from **base models** are passed to the **Voting Classifier**, which then makes the final decision.