Ensemble Models

- **Ensemble models** are machine learning models that combine predictions from multiple models to improve accuracy and robustness.

- The core idea behind ensemble learning is that combining multiple models can reduce the risk of errors associated with individual models, leading to better overall performance.

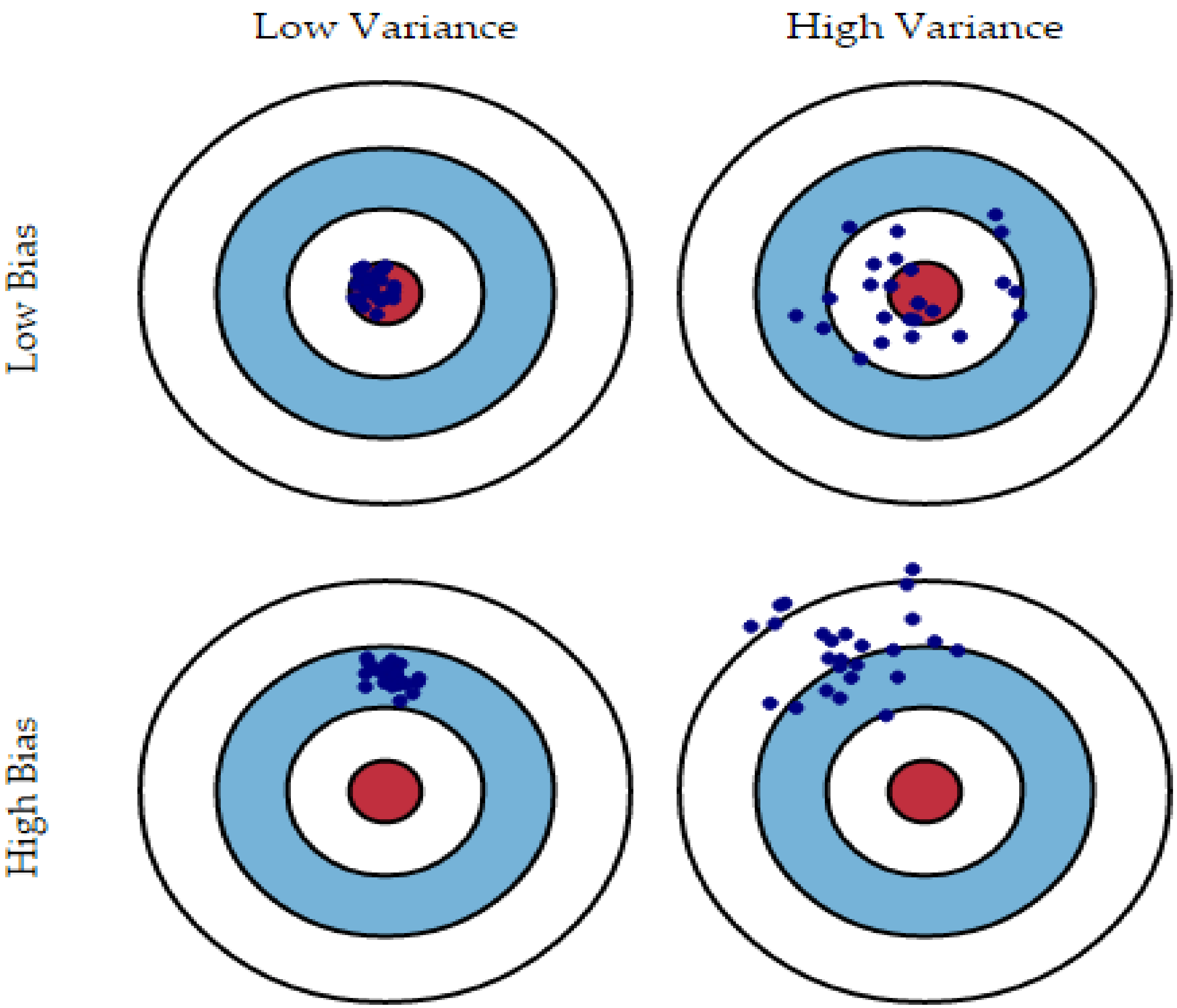
- **Ensemble methods** are widely used in both **classification** and **regression** tasks, and are especially useful when a single model does not perform well enough on its own.

Model Error
and
Reducing this Error with Ensembles

- The error emerging from any machine learning model can be broken down into three components mathematically:

Bias + Variance + Irreducible error

- **Bias error :**This is useful to quantify how much, on an average, the **predicted values are different from the actual value**.
- A high bias error means we have an underperforming model that keeps missing essential trends.
- Variance:Variance quantifies how the predictions made on the same observation differ.
- A high variance model will overfit your training population and perform poorly on any observation beyond training.
- The following diagram will give you more clarity (assume that the red spot is the real value, and the blue dots are predictions):
  :


Low Variance
High Variance
Low Bias
High Bias

- Typically, as you increase the complexity of your model, you will see a reduction in error due to lower bias in the model.

- However, this only happens until a particular point. As you continue to make your model more complex, you end up overfitting your model, and hence your model will start suffering from the high variance.

Types of Ensemble Models

There are different types of ensemble methods, and each one brings a set of advantages and disadvantages.

- Before diving into each method, let’s understand what **meta** and **base learners** are for a better understanding of the next concepts.
- **Base learners** are the first level of an ensemble learning architecture, and each one of them is trained to make individual predictions.
- **Meta learners**, on the other hand, are in the second level, and they are trained on the output of the base learners.

Key Types of Ensemble Methods

1. **Bagging (Bootstrap Aggregating)**
2. **Boosting**
3. **Stacking (Stacked Generalization)**
4. **Voting**

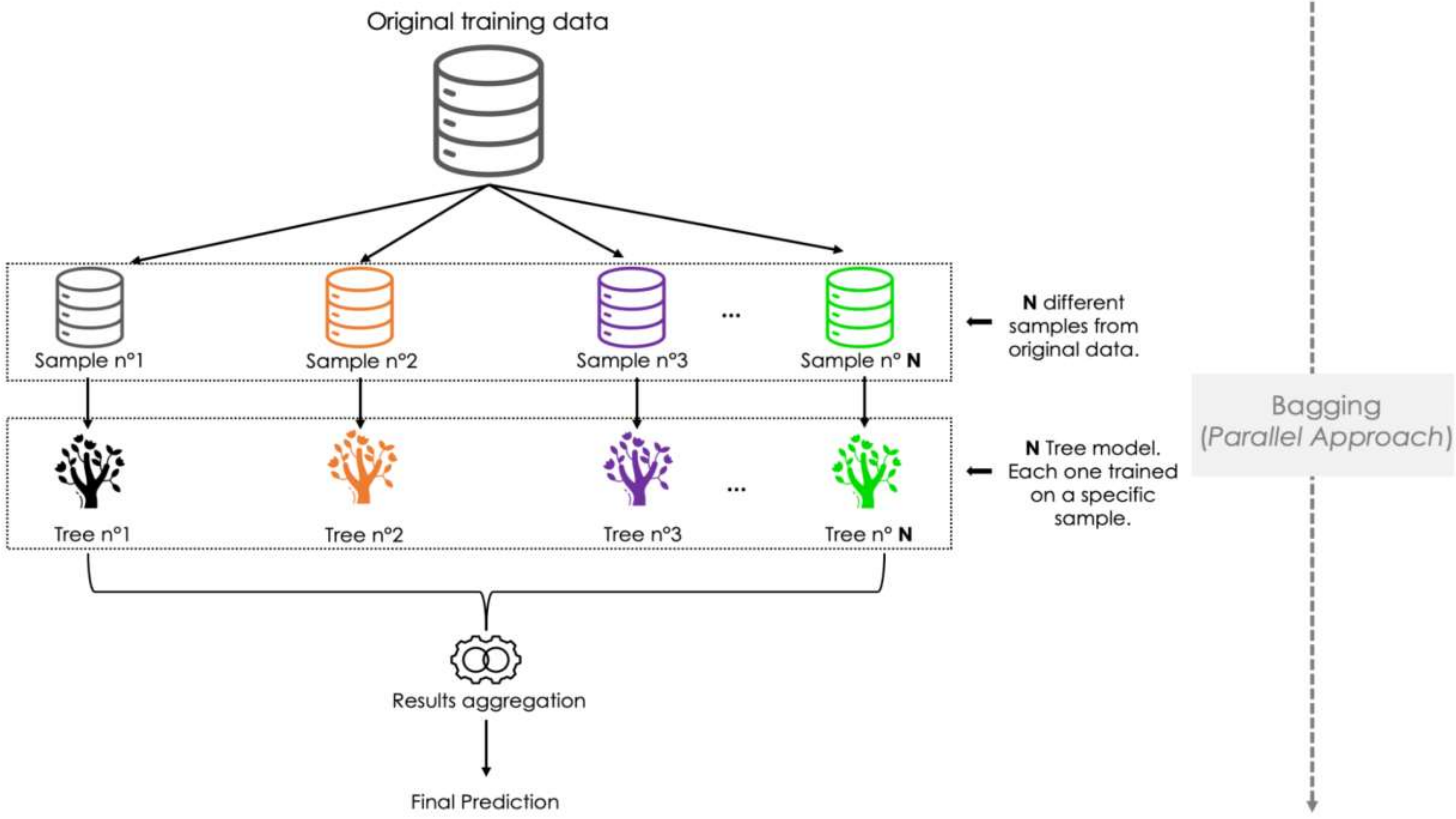
1. Bagging (Bootstrap Aggregating):

- **Bagging** is also known as **bootstrap aggregation**. This technique is similar to **random forest**, but it uses all the predictors, whereas random forest uses only a subset of predictors in each tree.

- In bagging, a random sample of data from the training set is selected with replacement - meaning that the individual data points can be chosen more than once.

• The main steps involved in bagging:

- ✓ Generation of multiple bootstrap resamples.
- ✓ Running an algorithm on each resample to make predictions.
- ✓ Combining the predictions by taking the average of the predictions or taking the majority vote (for classification).

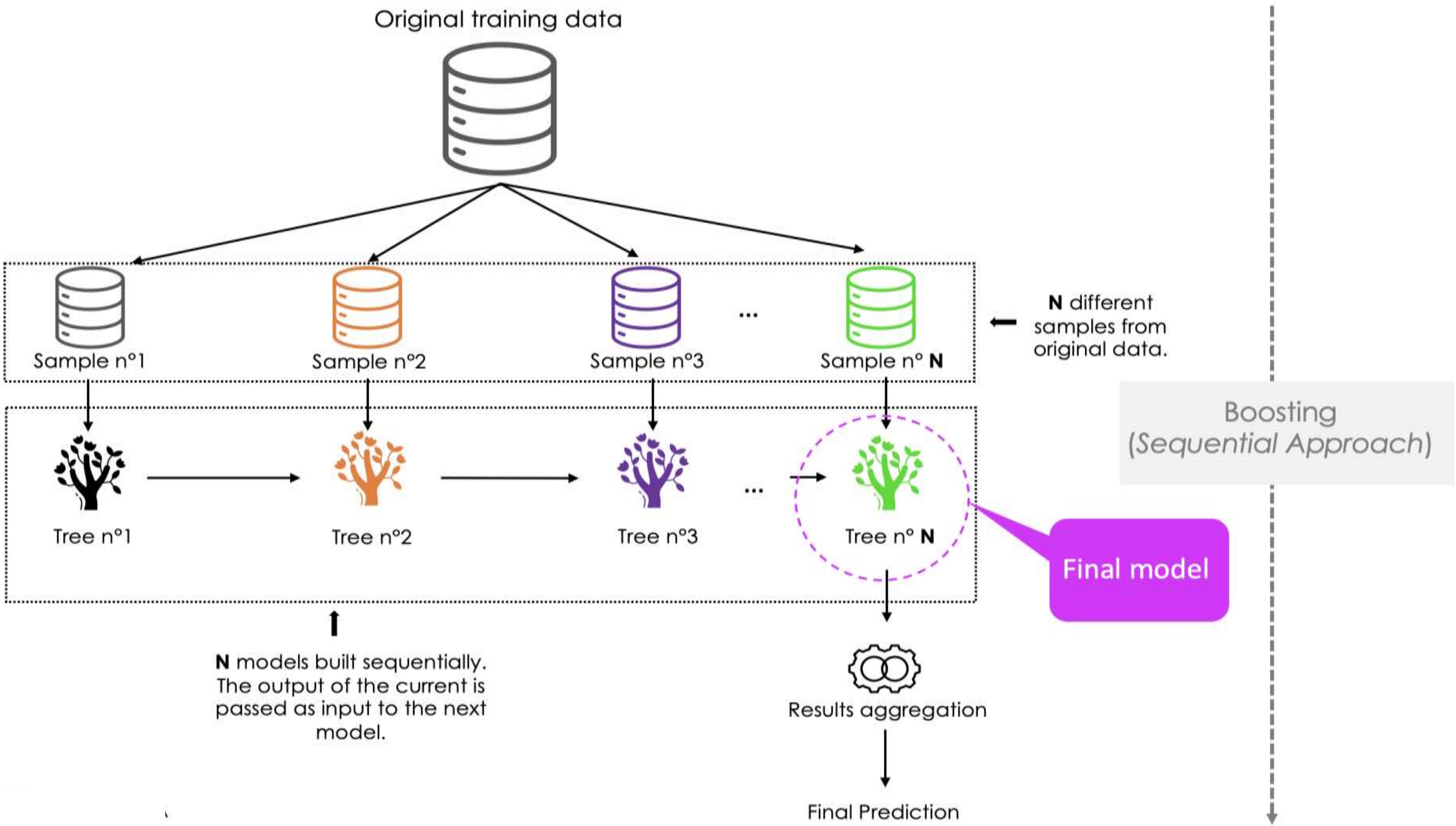

Original training data
Sample n°1
Sample n°2
Sample n°3
...
Sample n° N
N different samples from original data.
Tree n°1
Tree n°2
Tree n°3
...
Tree n° N
N Tree model. Each one trained on a specific sample.
Results aggregation
Final Prediction
Bagging
(Parallel Approach)

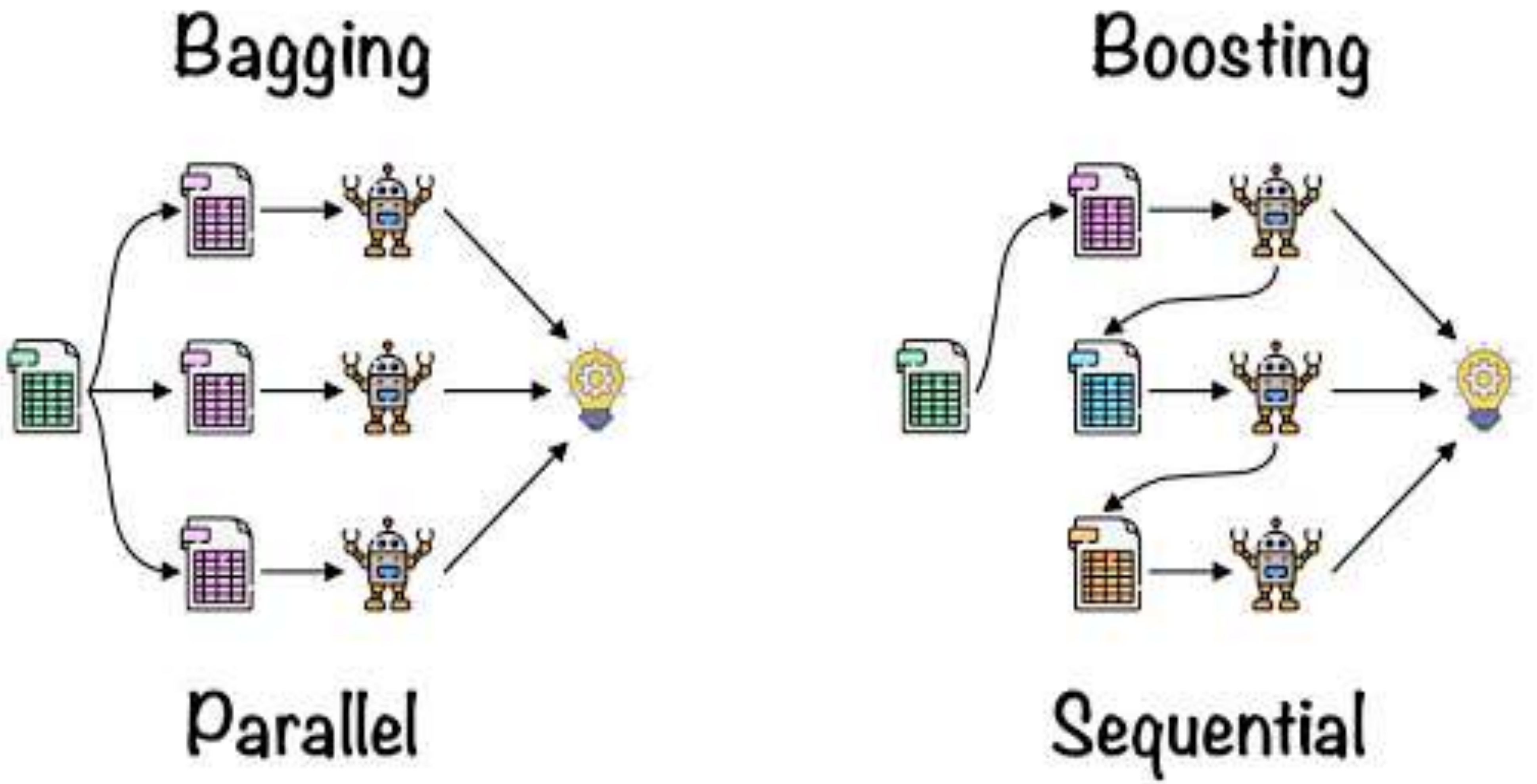
2. Boosting

- **Boosting** trains models sequentially, where each new model corrects the errors made by previous models.

- Each model iteratively focuses attention on the observations that are misclassified by its predecessors.

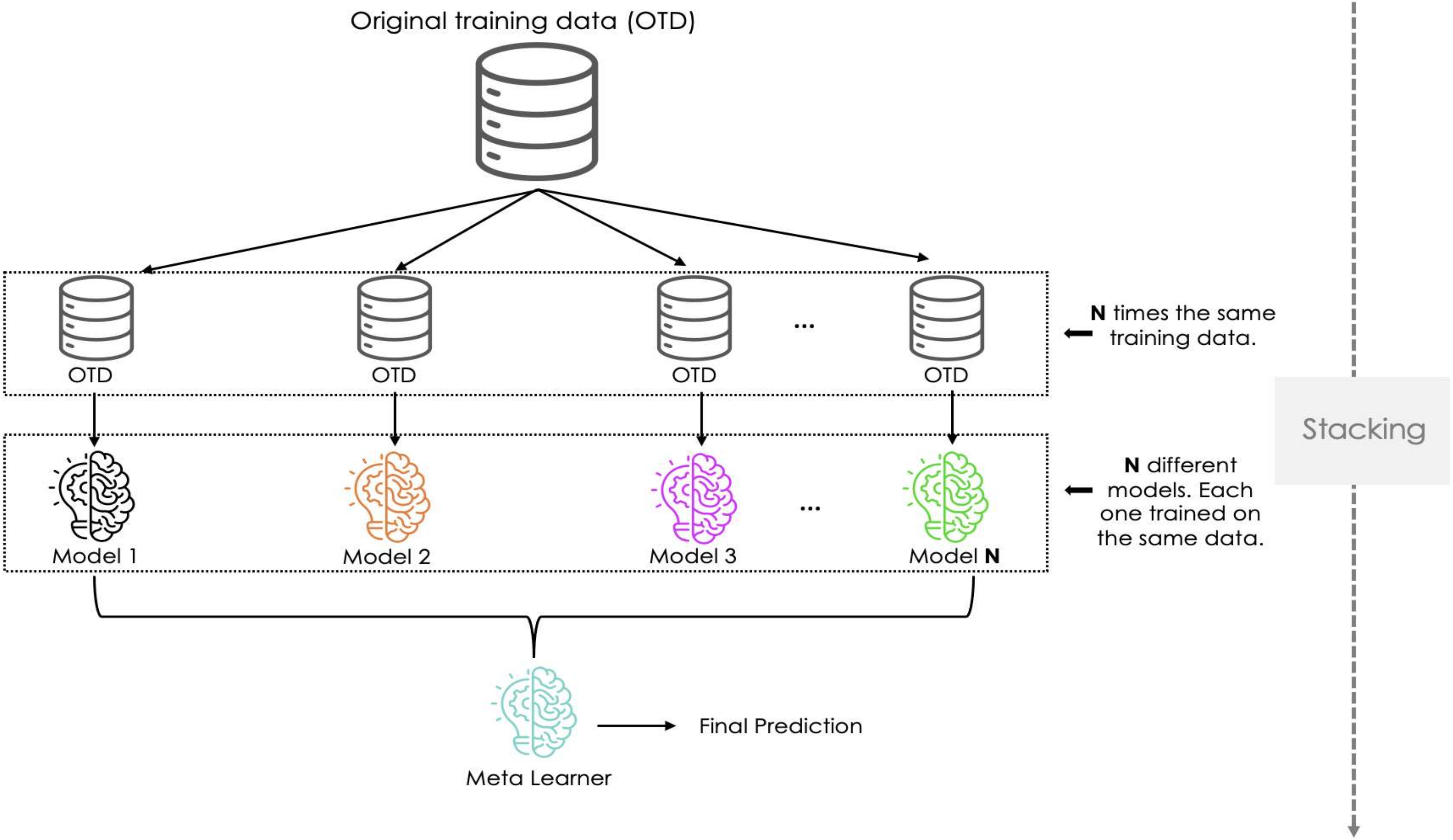
- **Examples**: AdaBoost, Gradient Boosting Machines (GBM), XGBoost, and LightGBM are popular boosting algorithms.

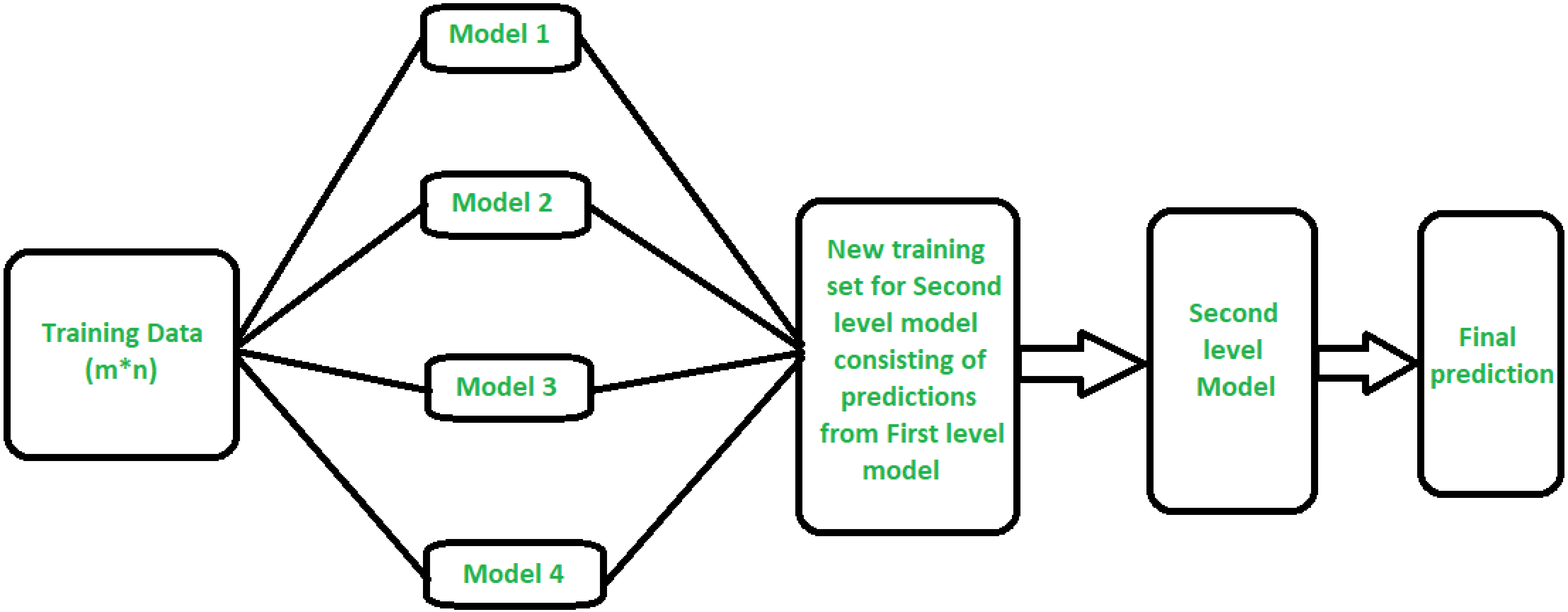

Original training data
Sample n°1
Sample n°2
Sample n°3
...
Sample n° **N**
N different samples from original data.
Tree n°1
Tree n°2
Tree n°3
...
Tree n° **N**
Final model
N models built sequentially. The output of the current is passed as input to the next model.
Results aggregation
Final Prediction
Boosting
(*Sequential Approach*)


Bagging
Boosting
Parallel
Sequential

3. Stacking (Stacked Generalization)

- The predictions from the base learners are stacked together and are used as the input to train the meta-learner to produce more robust predictions.
- **Stacking** is pretty similar to boosting.
- The meta learner is then used to make final predictions.
- Bagging and boosting typically use homogeneous base learners, whereas stacking tends to include heterogeneous ones.


Original training data (OTD)
OTD
OTD
OTD
...
OTD
N times the same training data.
Model 1
Model 2
Model 3
...
Model N
N different models. Each one trained on the same data.
Meta Learner
Final Prediction
Stacking

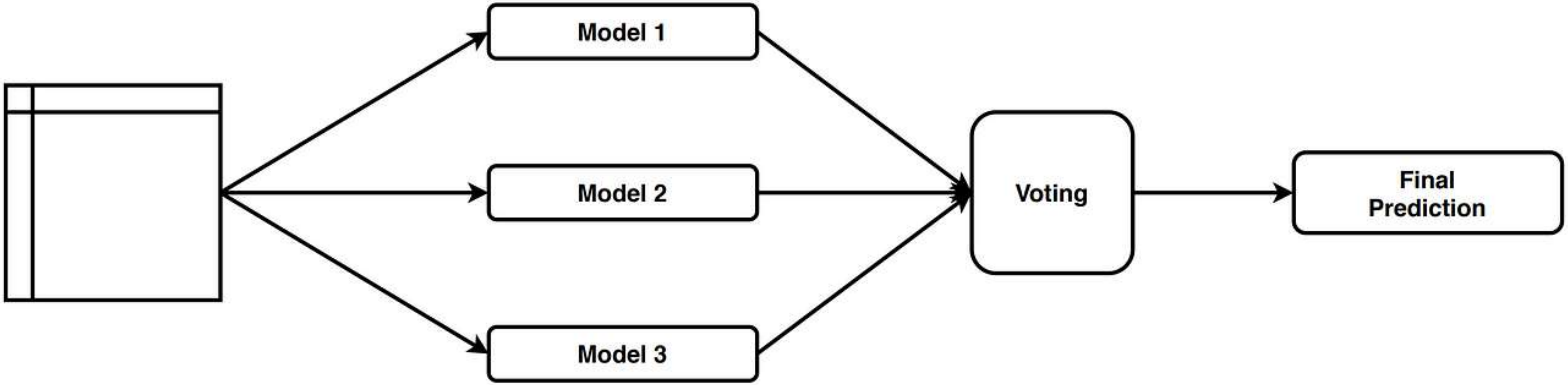

Training Data
(m*n)
Model 1
Model 2
Model 3
Model 4
New training set for Second level model consisting of predictions from First level model
Second level Model
Final prediction

4. Voting

- In **voting**, multiple models (often of different types, like logistic regression, decision trees, etc.) are trained on the same data.

- For classification tasks, the final prediction is made by majority voting (hard voting) or by averaging probabilities (soft voting).

- This method is simple and works best when the base models are diverse.


Model 1
Model 2
Model 3
Voting
Final
Prediction

Advantages of Ensemble Models

- **Improved Accuracy**: By combining multiple models, ensembles often outperform single models, especially on complex tasks.
- **Reduced Overfitting**: Techniques like bagging can reduce overfitting, as individual models are less likely to capture noise from the data.
- **Robustness**: Ensembles are generally more robust to errors and noise in data due to the diversity of models used.

Limitations of Ensemble Models

- **Increased Complexity**: Ensemble models are more complex and require more computational resources.

- **Interpretability**: With multiple models combined, interpreting the final model becomes challenging.

- **Longer Training Time**: Training multiple models or deep, complex models can be time-consuming.

Common Applications

- **Image and Text Classification**: Used in applications like facial recognition, text sentiment analysis, etc.

- **Medical Diagnosis**: Helps to combine insights from different types of data (e.g., imaging, lab results) for diagnosis.

- **Financial Forecasting**: Used in stock price prediction, risk assessment, and credit scoring.

- **Recommendation Systems:** Commonly used to predict user preferences based on various model types (e.g., collaborative filtering, content-based).

- **Ensemble models** are among the most powerful techniques in machine learning, balancing the biases and variances of individual models to improve predictive performance.