Predicting the price of a house using machine learning

Ponjesly College Of Engineering

Department Of Computer Science And Engineering

Submitted By
Amaresh

Ahilesh

Abish

Altrin

Gopi Krishna

Predicting the price of a house using machine learning is a complex task that can be affected by several drawbacks and challenges. Overcoming these drawbacks requires a systematic approach and the application of various techniques. Here's a full topic breakdown on how to address the drawbacks in predicting house prices using machine learning:

1. Data Quality and Quantity
2. Feature Engineering
3. Model Selection
4. Overfitting
5. Data Scaling and Normalization
6. Hyperparameter Tuning
7. Evaluation Metrics

1. Data Quality and Quantity:
   - Drawback: Limited or noisy data can hinder model accuracy.
   - Solution:
     - Collect a diverse and comprehensive dataset.
     - Clean the data by handling missing values and outliers.
     - Augment data if necessary through techniques like data synthesis or data imputation.

2. Feature Engineering:
- Drawback: Insufficient or irrelevant features can result in poor predictions.
- Solution:
    - Carefully select relevant features using domain knowledge.
    - Create new features through feature engineering to capture hidden patterns.
    - Use techniques like PCA (Principal Component Analysis) for dimensionality reduction
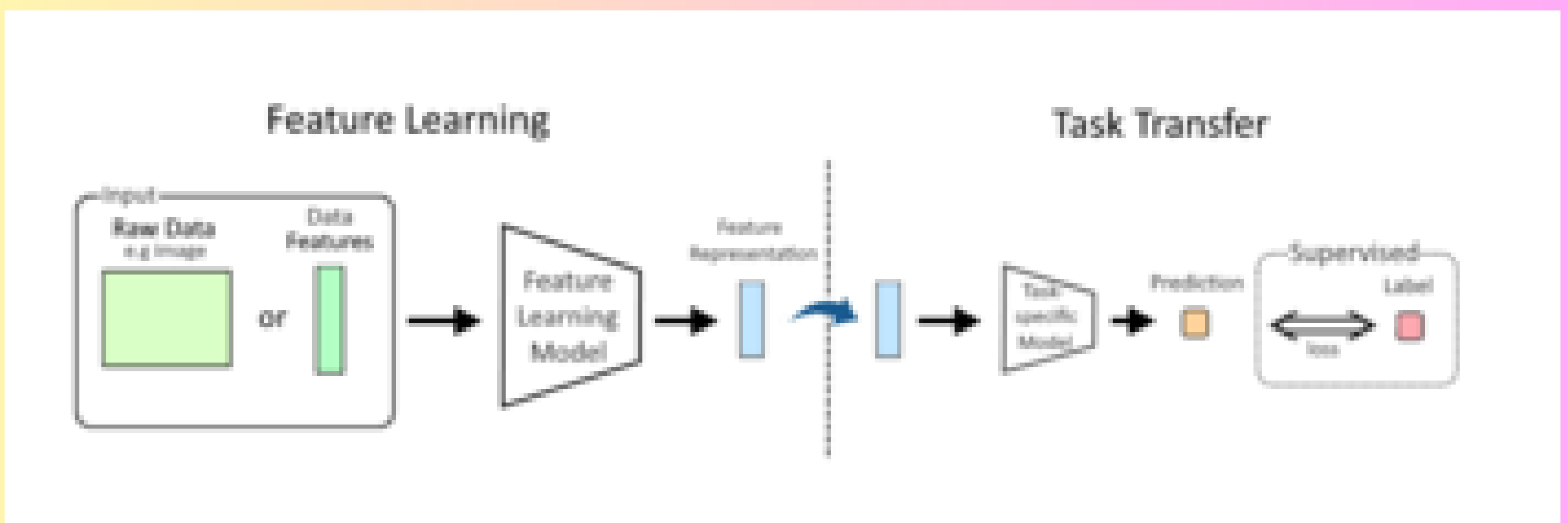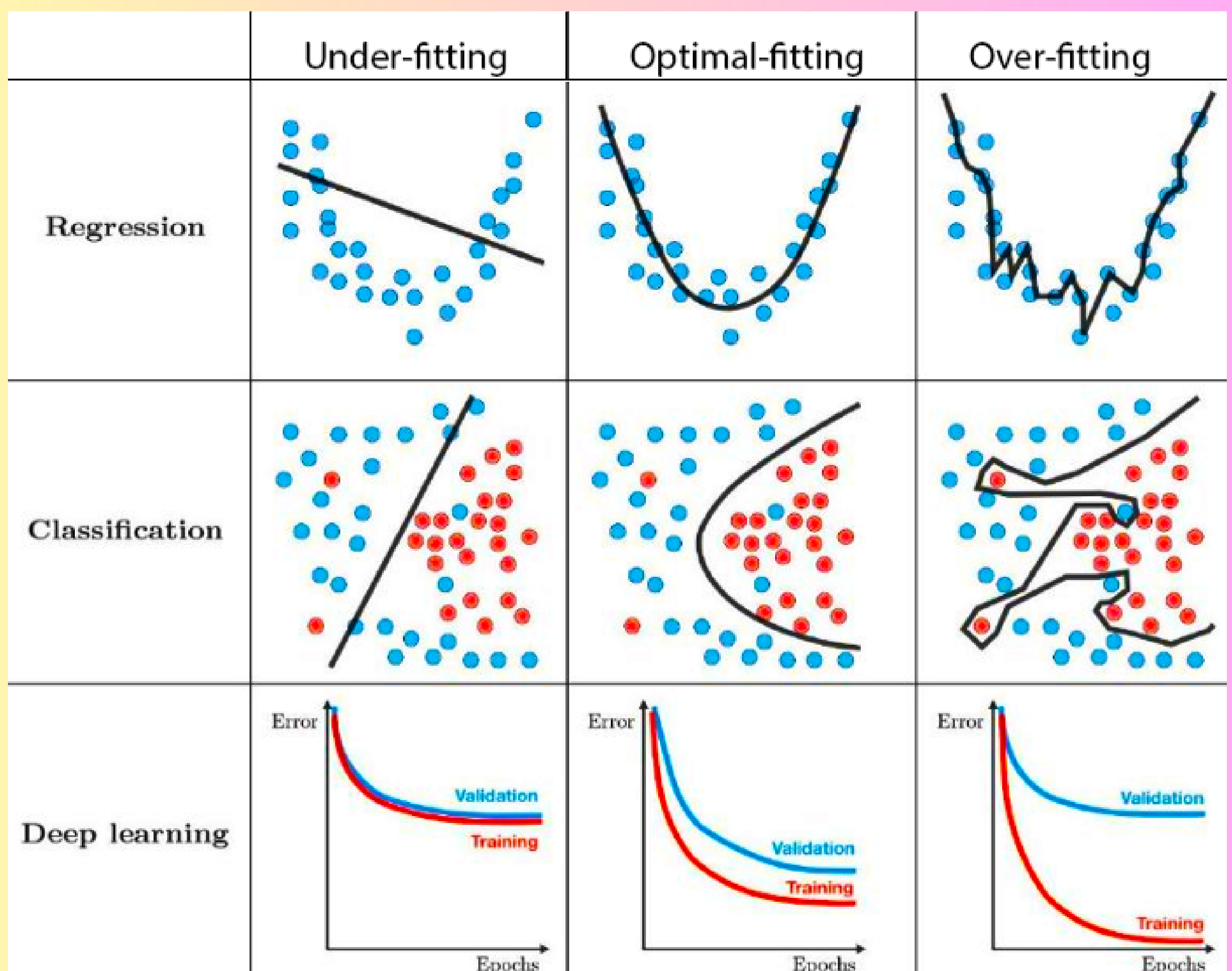
3. Model Selection:

- Drawback: Choosing the wrong algorithm or model architecture can lead to poor performance.
- Solution:
  - Experiment with various algorithms, such as linear regression, decision trees, random forests, support vector machines, and neural networks.
  - Use techniques like cross-validation to evaluate model performance and select the best-performing model.
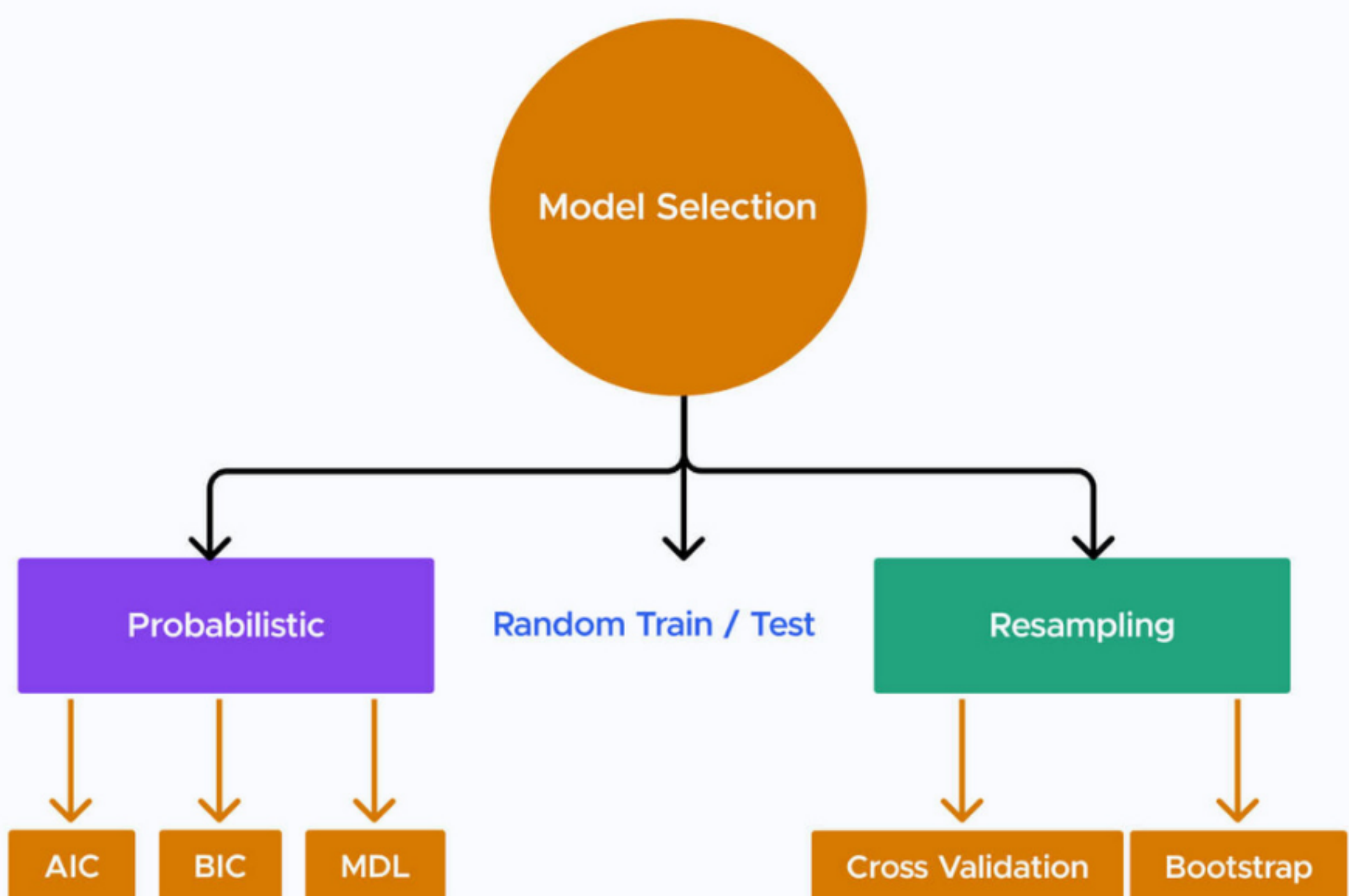
4. Overfitting:

- Drawback: Models may perform well on training data but poorly on new data due to overfitting.
- Solution:
    - Regularize models using techniques like L1 or L2 regularization.
    - Split data into training, validation, and test sets to monitor overfitting.
    - Use ensemble methods like bagging and boosting to reduce overfitting.

|	Under-fitting	Optimal-fitting	Over-fitting
Regression			
Classification			
Deep learning	Error / Validation / Training / Epochs	Error / Validation / Training / Epochs	Error / Validation / Training / Epochs

5. Data Scaling and Normalization:

- Drawback: Features with different scales can affect model convergence.
- Solution:
  - Standardize or normalize features to have a consistent scale.
  - Use techniques like Min-Max scaling or Z-score normalization.
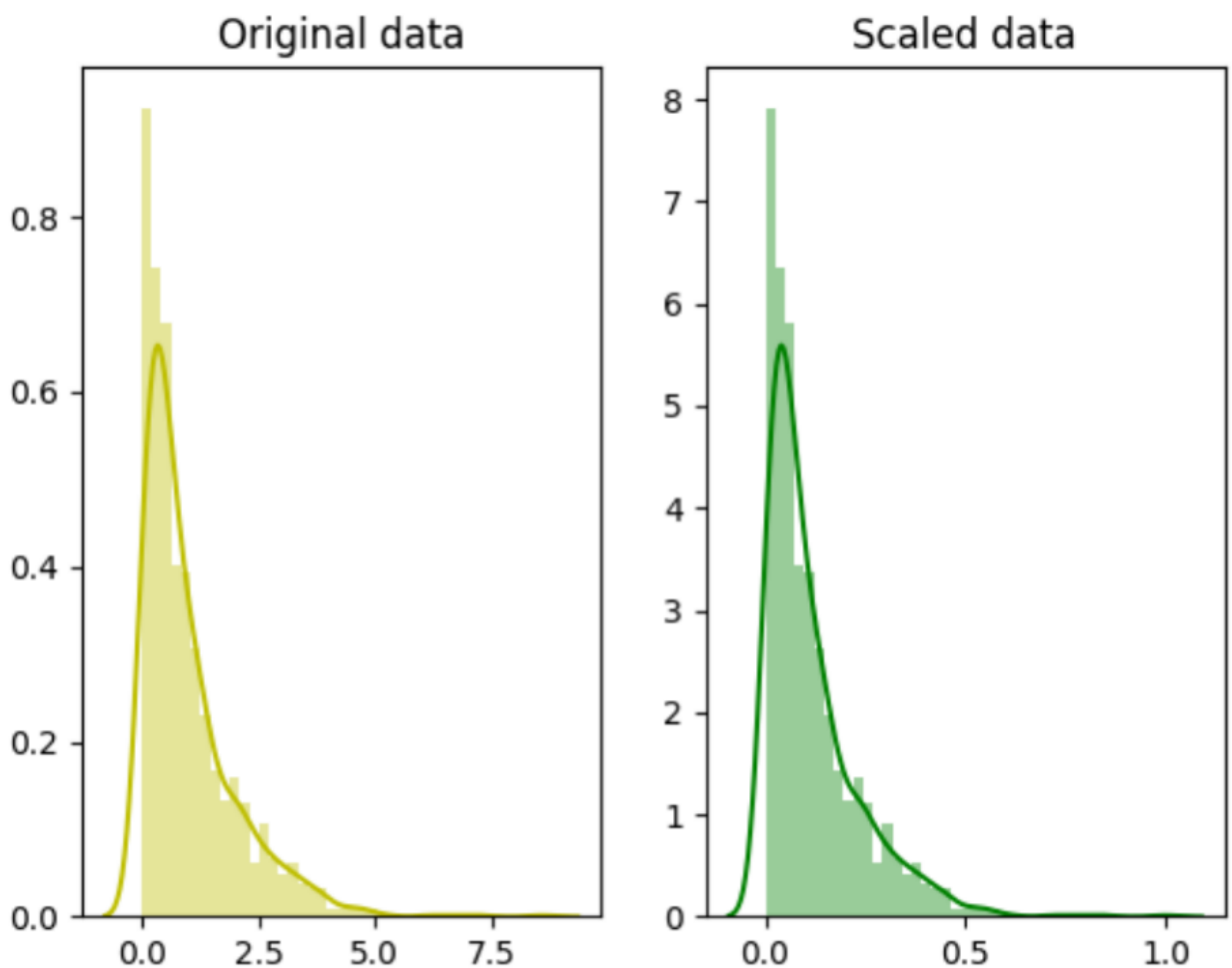
6. Hyperparameter Tuning:

- Drawback: Poorly tuned hyperparameters can result in suboptimal model performance.

- Solution:
    - Utilize techniques like grid search or random search to find the best hyperparameters.
    - Consider using automated hyperparameter optimization tools like Bayesian optimization.
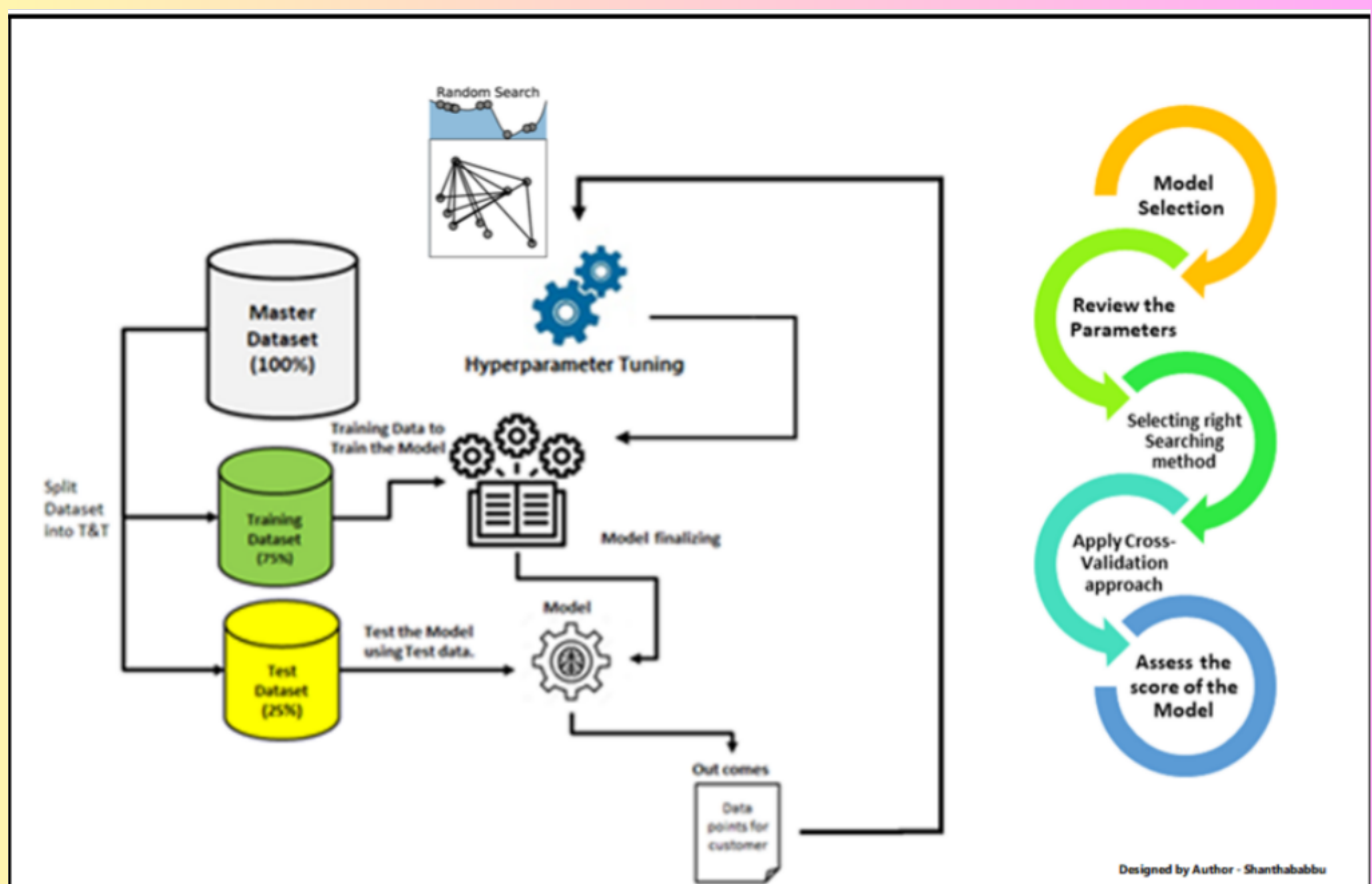
7. Evaluation Metrics:

- Drawback: Using the wrong evaluation metrics can misrepresent model performance.
- Solution:
  - Choose appropriate evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE) for regression tasks.
  - Consider domain-specific metrics or business objectives when applicable.

Metric	Formula
True positive rate, recall	$\dfrac{TP}{TP+FN}$
False positive rate	$\dfrac{FP}{FP+TN}$
Precision	$\dfrac{TP}{TP+FP}$
Accuracy	$\dfrac{TP+TN}{TP+TN+FP+FN}$
F-measure	$\dfrac{2 \cdot precision \cdot recall}{precision + recall}$