Data Analytics

Model Evaluation Metrics

Understanding Key Performance Indicators

Classification • Regression • Clustering

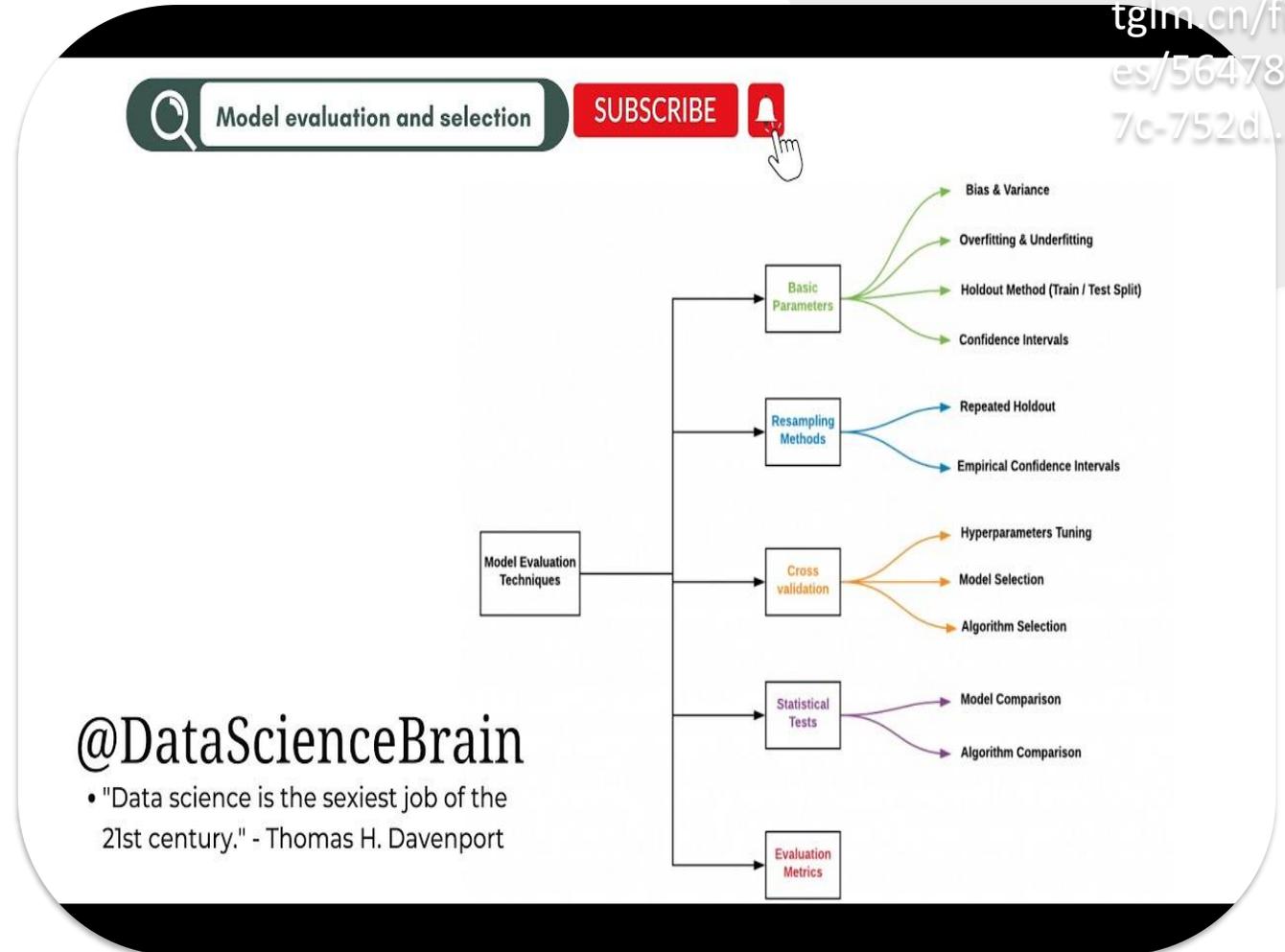# Introduction to Model Evaluation Metrics

! **Why Evaluation Matters**

- Validate model performance

- Measure prediction accuracy

- Identify model weaknesses

- Ensure business goals met

✔ **Key Benefits**

- Improved decision-making

- Optimize model selection

Model Evaluation & Selection Process

Types of Evaluation Metrics

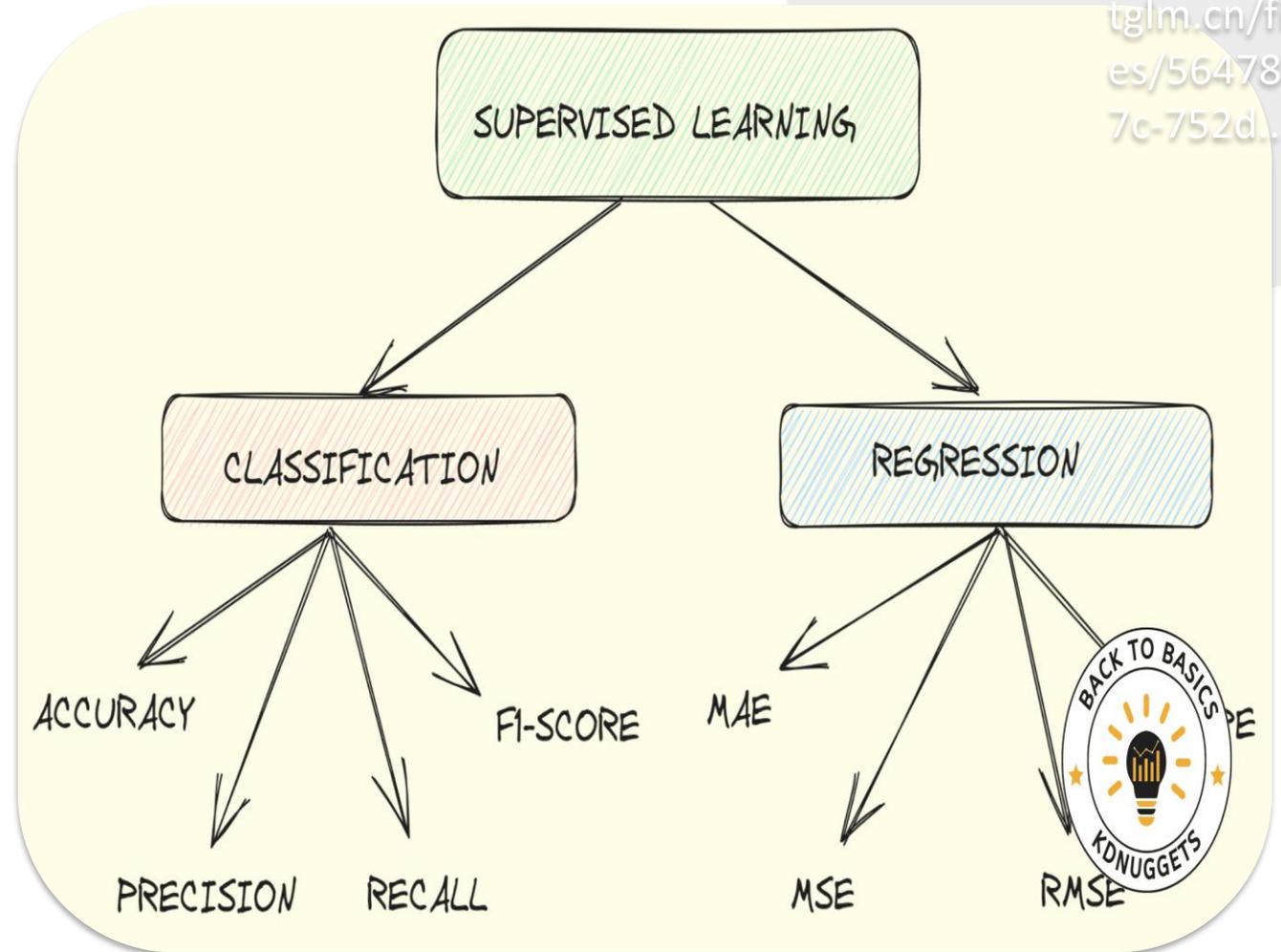Three main categories based on ML task type

▲● Classification

Predict discrete class labels

〰 Regression

Predict continuous values

○ Clustering

Group similar data points

Supervised Learning: Classification vs. Regression

Classification Metrics - Part 1

Foundation metrics for binary classification

⊞ Confusion Matrix

2x2 table of predictions vs. actuals

TP — **True Positive**
Correctly predicted positive

TN — **True Negative**
Correctly predicted negative

FP — **False Positive**
Incorrectly predicted positive

FN — **False Negative**
Incorrectly predicted negative

Confusion Matrix Structure and Components

Classification Metrics - Part 2

Advanced performance metrics
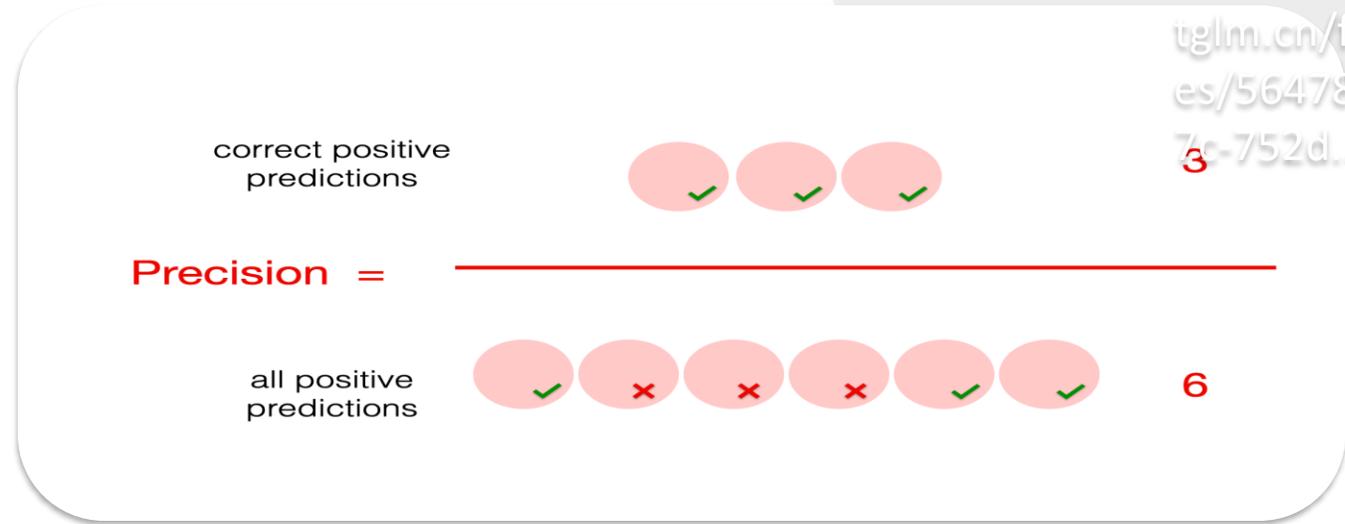
Precision

TP / (TP + FP)
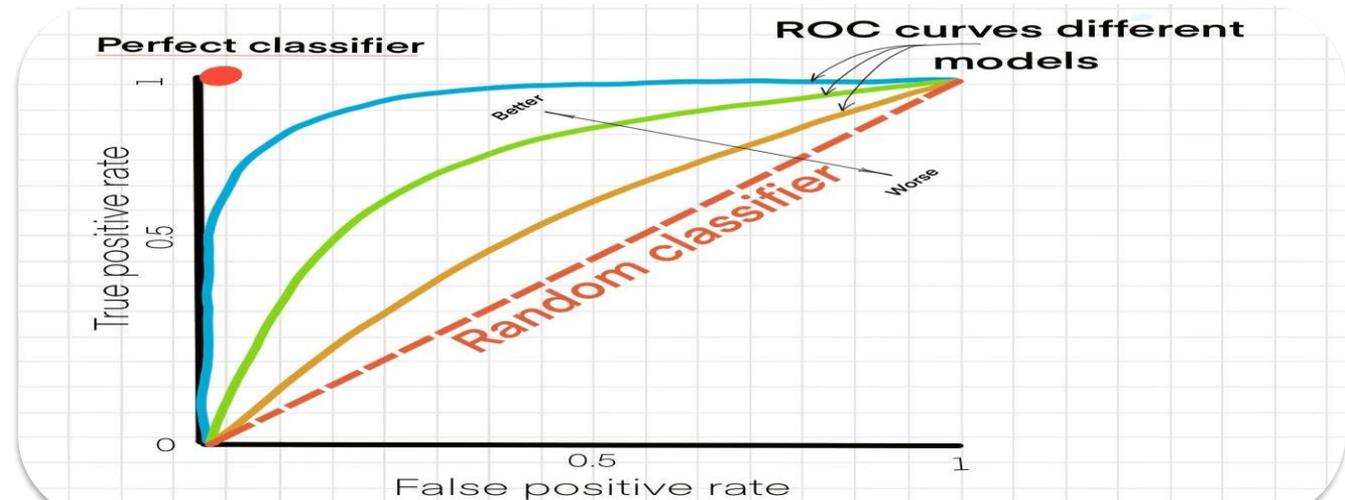
Recall / Sensitivity

TP / (TP + FN)

F1-Score

2 × (Precision × Recall) / (Precision + Recall)

ROC / AUC

True Positive Rate vs. False Positive Rate

correct positive predictions

Precision =

all positive predictions

3

6

Precision Calculation

Perfect classifier

ROC curves different models

True positive rate

Better

Worse

Random classifier

False positive rate

Regression Metrics

Evaluating continuous predictions

Σ MSE

Mean Squared Error

RMSE

Root Mean Squared Error

MAE

Mean Absolute Error

R²

Coefficient of Determination

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

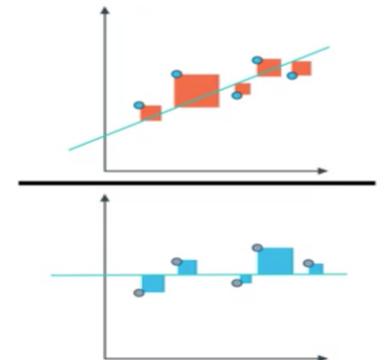Predicted Value vs. Actual Value
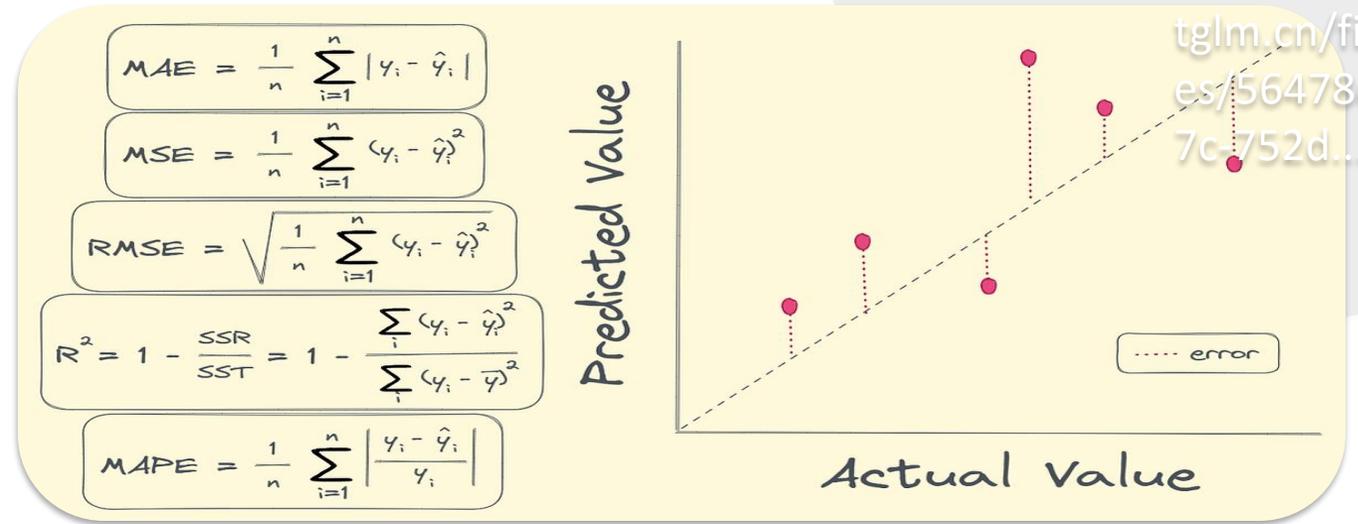
..... error

Predicted vs. Actual Values with Errors

○ R2 SCORE

BAD MODEL
The errors should be similar.
R2 score should be close to 0.

GOOD MODEL
The mean squared error for the linear regression model should be a lot smaller than the mean squared error for the simple model.
R2 score should be close to 1.

$$R2 = 1 -$$

Cross-Validation Techniques

Robust model performance assessment

K-Fold CV

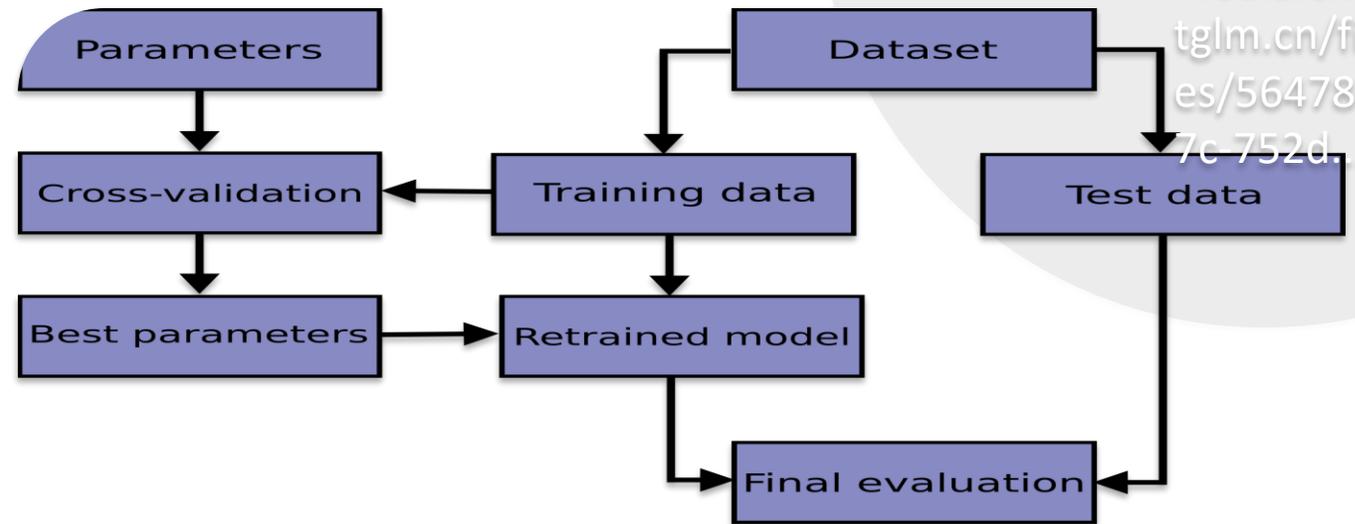Split data into K equal folds

Stratified K-Fold

Preserves class distribution

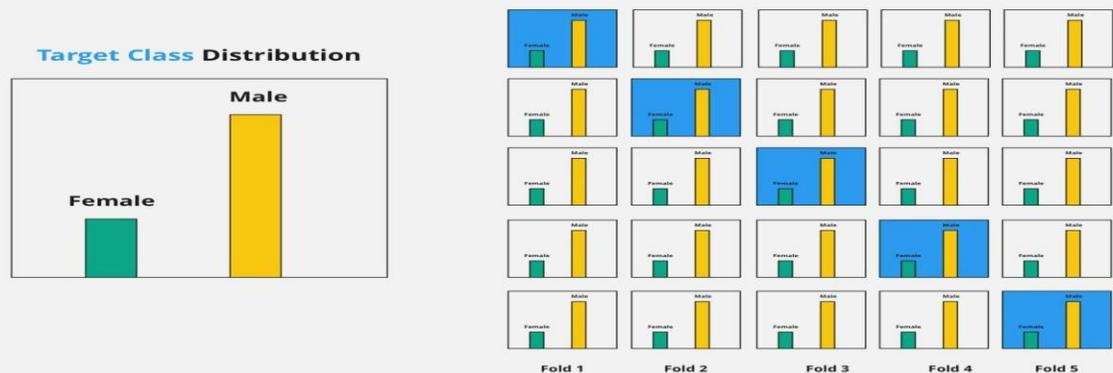**Leave-One-Out**

Each sample as test set once

Benefits

Reduces overfitting risk

Cross-Validation Workflow

Bias-Variance Tradeoff

Balancing model complexity

↘ Underfitting

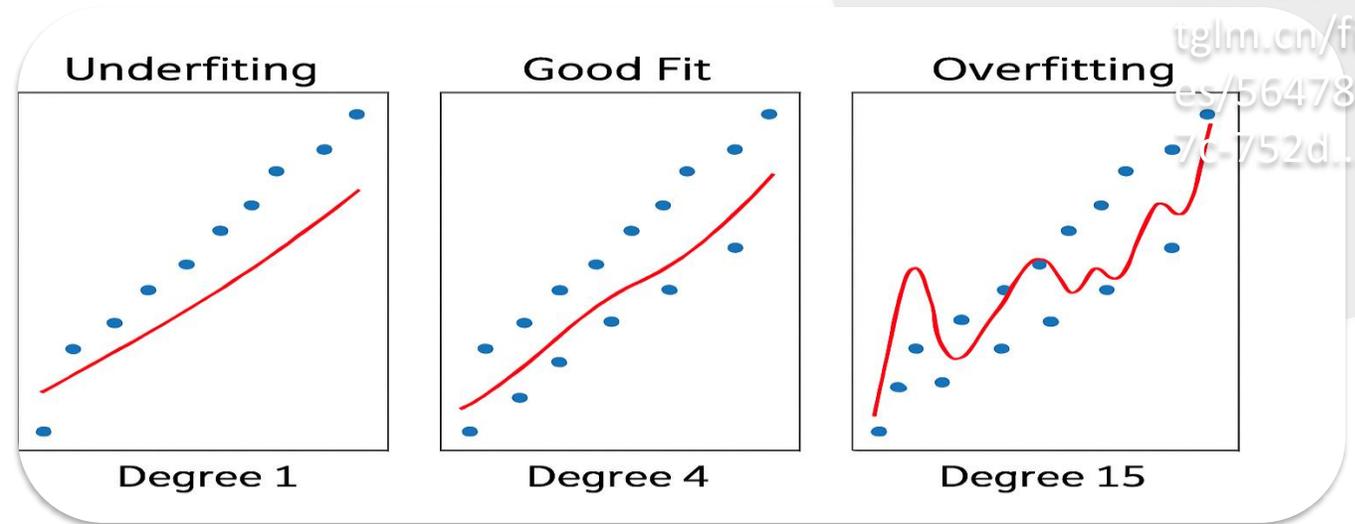High bias, low variance

↗ Overfitting
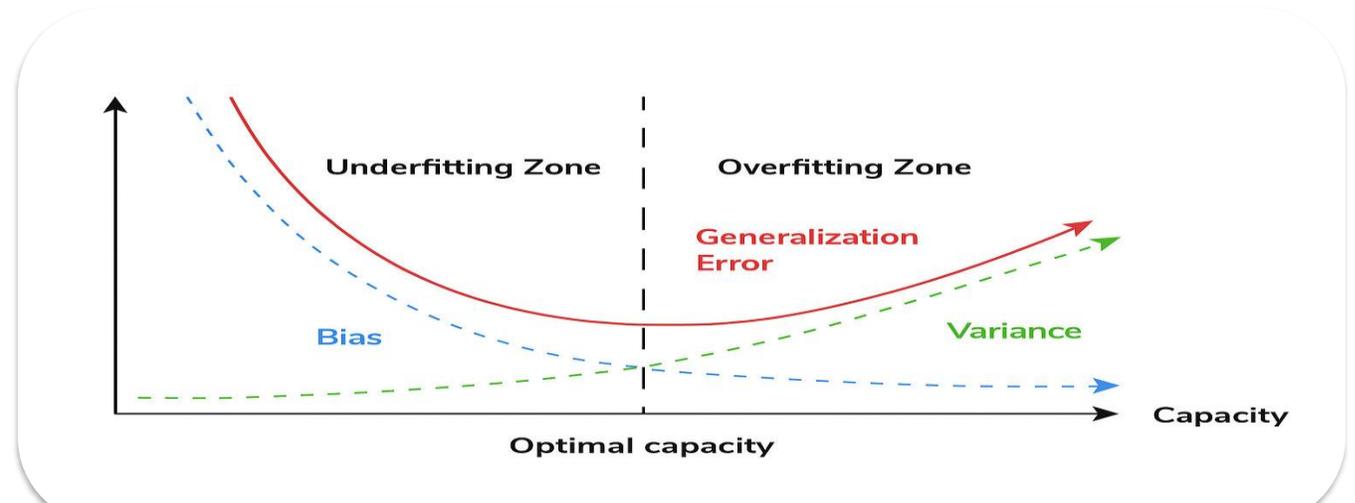
Low bias, high variance

✓ Optimal Fit

Balanced bias & variance

⚑ Goal

Minimize total error

Underfitting, Good Fit, and Overfitting Comparison

Mind Map - Model Evaluation Metrics

Comprehensive Overview

GenAI-generated mind map organizing all evaluation metrics
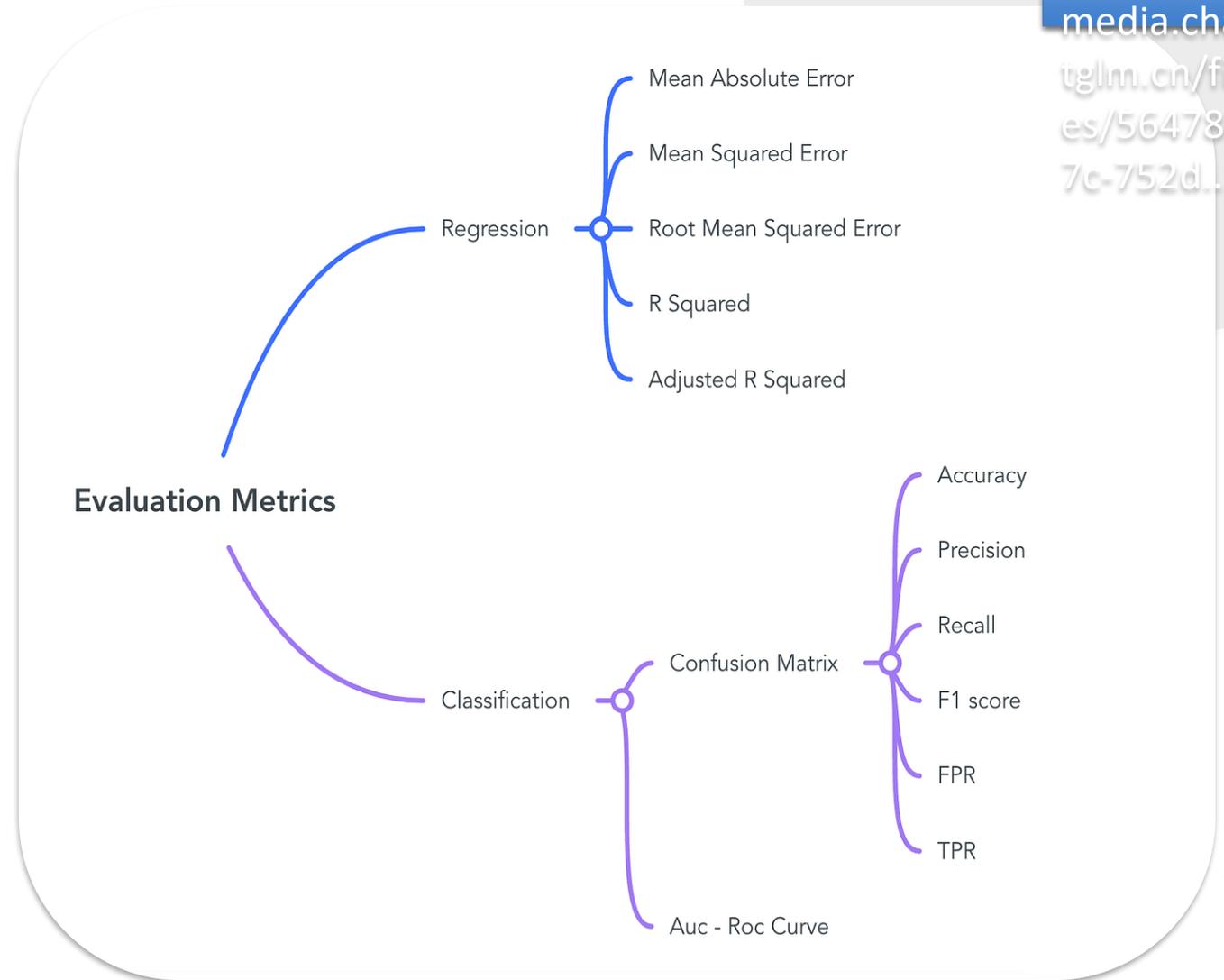
▲ Classification Metrics

Accuracy, Precision, Recall, F1, ROC/AUC

〜 Regression Metrics

MSE, RMSE, MAE, R²

⠐ Clustering Metrics

Silhouette Score, Davies-Bouldin Index

Evaluation Metrics

- Regression
  - Mean Absolute Error
  - Mean Squared Error
  - Root Mean Squared Error
  - R Squared
  - Adjusted R Squared
- Classification
  - Confusion Matrix
    - Accuracy
    - Precision
    - Recall
    - F1 score
    - FPR
    - TPR
  - Auc - Roc Curve

Summary and Recap

Key Learning Points

Classification

Confusion Matrix, Accuracy, Precision, Recall, F1, ROC/AUC

Regression

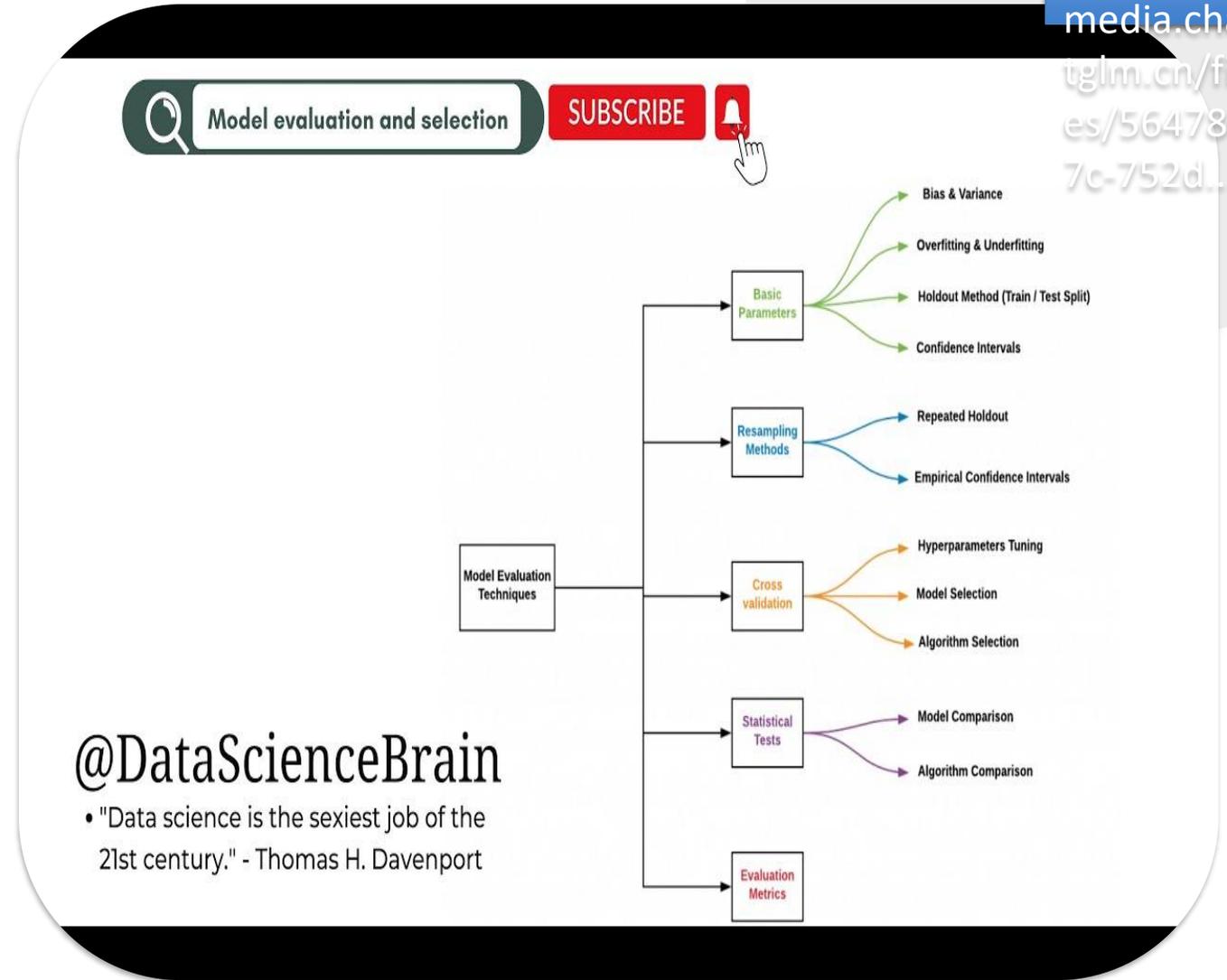MSE, RMSE, MAE, R²

Cross-Validation

K-Fold, Stratified K-Fold, Leave-One-Out

Bias-Variance Tradeoff

Balance underfitting & overfitting

Critical Importance

Proper evaluation ensures reliable models