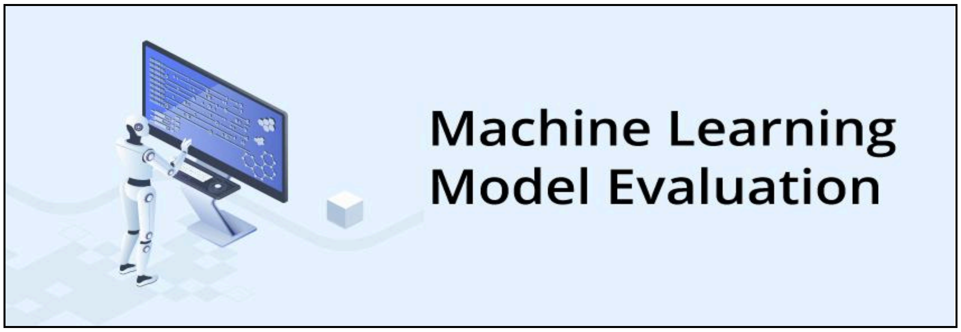

39/100
31 JAN 24| DAY - 39| MACHINE LEARNING
#100DAYSOFDATA SCIENCE
PYTHON | NUMPY | PANDAS | MATPLOTLIB | SEABORN | SQL | STATS | MACHINE LEARNING |



Model Evaluation Techniques

---

1. Accuracy:

Accuracy measures the ratio of correctly predicted instances to the total instances. It's suitable for balanced datasets but can be misleading for imbalanced ones.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

2. Precision:

Precision measures the proportion of correctly predicted positive instances out of all instances predicted as positive. It's important when false positives are costly.

$$Precision = \frac{TP}{TP + FP}$$

3. Recall (Sensitivity or True Positive Rate):

Recall measures the proportion of correctly predicted positive instances out of all actual positive instances. It's important when false negatives are costly.

$$Recall = \frac{TP}{TP + FN}$$

4. F1-Score:

F1-score is the harmonic mean of precision and recall. It balances precision and recall, making it useful when you want to consider both false positives and false negatives.

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall}$$

5. Specificity (True Negative Rate):

Specificity measures the proportion of correctly predicted negative instances out of all actual negative instances.

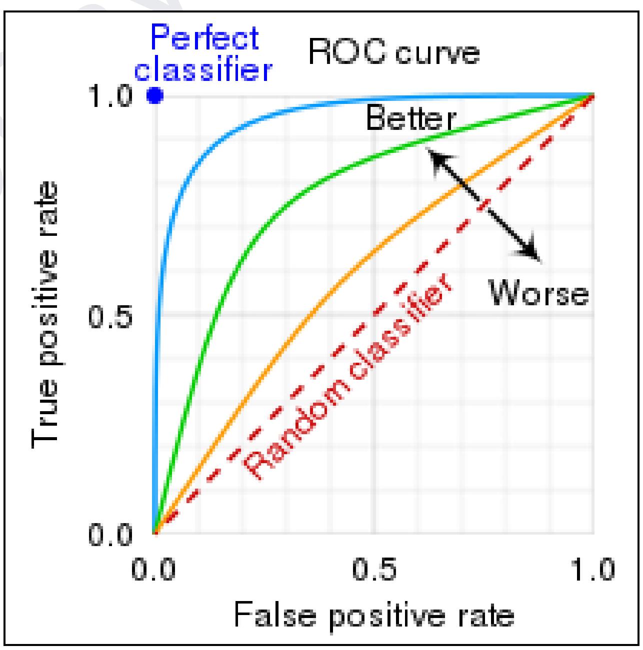
$$\textbf{Specificity} = \frac{\textbf{True Negatives}}{\textbf{True Negatives} + \textbf{False Positives}}$$

6. ROC Curve (Receiver Operating Characteristic):

The ROC curve visualizes the tradeoff between true positive rate (recall) and false positive rate as the classification threshold changes.

7. AUC (Area Under the ROC Curve):

AUC quantifies the area under the ROC curve. It's a single value that measures the model's ability to distinguish between classes.



The True Positive Rate or Recall is defined as

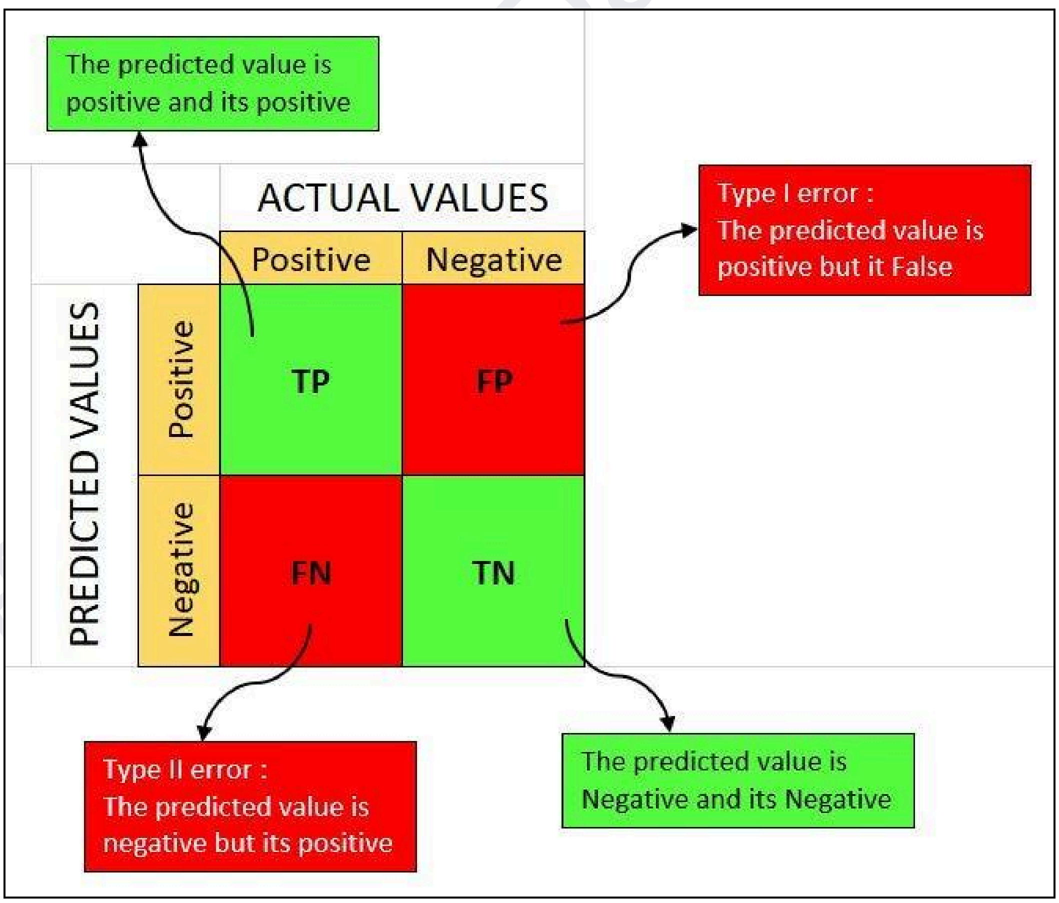
$$TPR = \frac{TP}{TP + FN}$$

The False Positive Rate is defined as

$$FPR = \frac{FP}{FP + TN}$$

8. Confusion Matrix:

A confusion matrix displays the counts of true positives, true negatives, false positives, and false negatives. It's useful for understanding model performance in classification tasks.



9. Mean Squared Error (MSE):

MSE measures the average squared difference between predicted and actual values in regression tasks. It quantifies the model's predictive accuracy.

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

10. Root Mean Squared Error (RMSE):

RMSE is the square root of MSE. It's more interpretable as it's in the same unit as the target variable.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

11. R-squared (Coefficient of Determination):

R-squared measures the proportion of the variance in the dependent variable that's explained by the model. It assesses the model's goodness of fit.

$$\mathrm{R}^2 = 1 - \frac{\text{Unexplained Variation}}{\text{Total Variation}}$$

12. Mean Absolute Error (MAE):

MAE calculates the average absolute difference between predicted and actual values. It's another metric for evaluating regression models.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

13. Log-Loss:

Log loss, aka logistic loss or cross-entropy loss. A measure of the accuracy of a probabilistic classifier's predictions.

$$L_{\log}(y, p) = -(y \log(p) + (1 - y) \log(1 - p))$$

14. Learning Curves:

Learning curves show how model performance changes with the size of the training data. They help diagnose underfitting or overfitting.

$$\mathbf{Y} = \mathbf{a}\mathbf{X}^{\mathrm{b}}$$

15. Feature Importance Analysis:

Identifies the importance of each feature in influencing model predictions. Useful for understanding feature relevance and potential model improvements.

16. Residual Analysis:

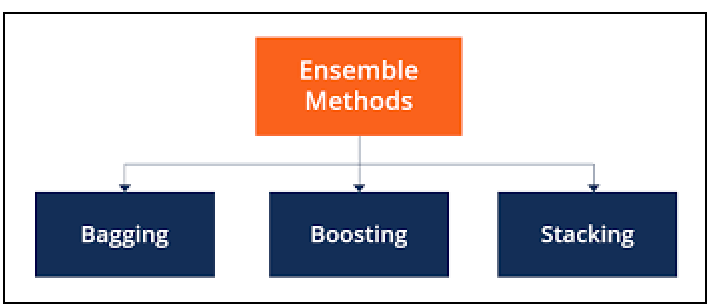
In regression tasks, residual analysis examines the patterns in the model's errors (residuals) to check for any systematic deviations.

17. Hyperparameter Tuning:

Adjusting hyperparameters to optimize model performance. Techniques include grid search, random search, and Bayesian optimization.

18. Ensemble Methods:

Using combinations of multiple models to improve overall predictive performance. Techniques include bagging, boosting, and stacking.



19. Domain-Specific Metrics:

Some domains have specialized evaluation metrics. For example, Mean Average Precision (MAP) in information retrieval or Gini coefficient in economics.

20. Cross-Validation:

A technique for assessing a model's performance by splitting data into multiple subsets, training and testing on different subsets.