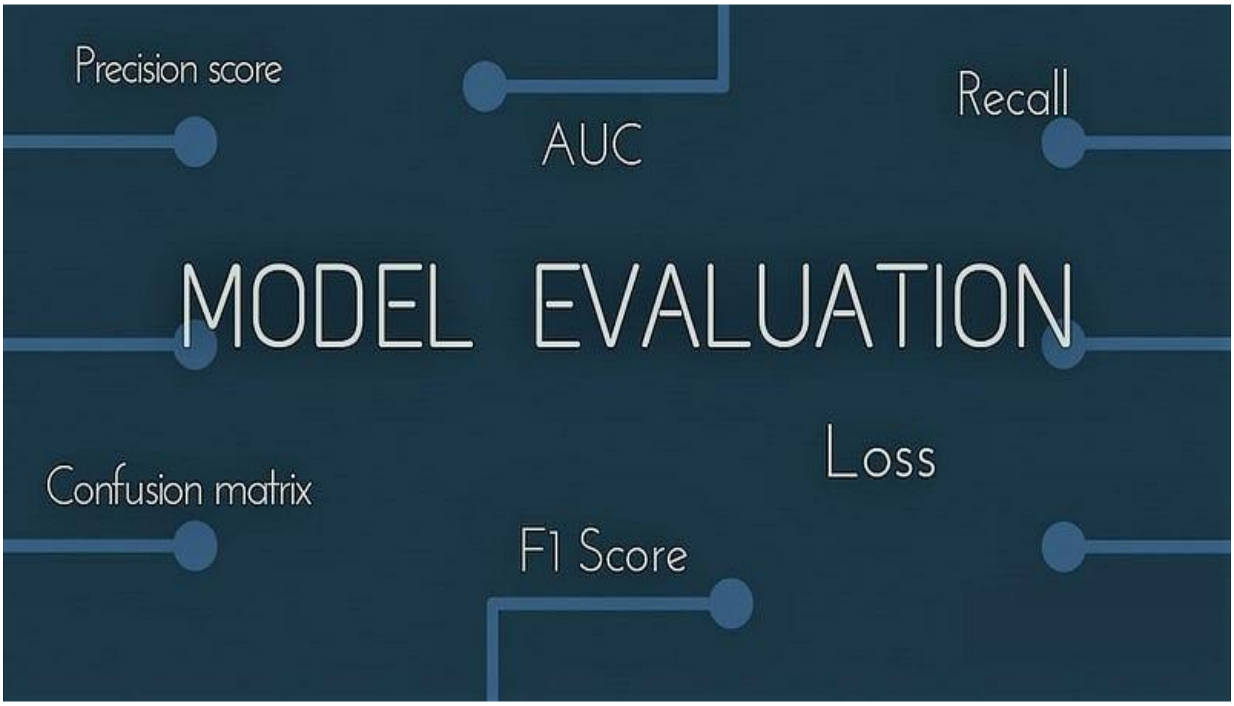
MENTORNESS ARTICLE

TASK 1

MIP-ML-08 BATCH

Evaluation Metrics in Machine Learning: Exploring Performance Assessment

BY NIGAR SULTANA



What Are Evaluation Metrics in Machine Learning?

Evaluation metrics in Machine Learning (ML) are important tools for checking how well ML models work. These metrics use numbers to show how effective a model is at handling new information and help decide which model is best. They look at things like accuracy, precision, recall, and other measures to see if the model is performing well. The goal is to make sure the model can learn from training and make accurate predictions on new things it hasn't seen before. These metrics also help pick the best model from many options and can find where a model needs to improve. This helps ML experts adjust things to make the model more effective and ensures it does well in real-life situations.

Why Do We Need Evaluation Metrics in ML?

Evaluation metrics are crucial in ML for several reasons:

- They help us understand if ML models are effective and accurate in their predictions or classifications.
- These metrics guide us in choosing the best model among several options by comparing their performances.
- They play a role in tuning model settings (hyperparameters) to make them perform better.
- Evaluation metrics provide a foundation for constantly improving and fine-tuning ML algorithms.
- They allow us to objectively assess and compare different models based on their performance scores.
- These metrics are essential for making informed decisions about which ML model to use in a given situation.
- They help us identify areas where ML models need improvement or adjustments.
- Using evaluation metrics ensures that ML models meet desired performance standards and deliver reliable results.
- Overall, evaluation metrics are key tools for assessing, improving, and optimizing ML models for various applications.

Types of Evaluation Metrics

There are various types of evaluation metrics used in ML, including:

1. **Regression Model Evaluation Metrics:** These metrics assess how well regression models predict numerical outcomes. They include:

    - **Mean Absolute Error (MAE):** This measures the average difference between predicted and actual values, giving an idea of how accurate the predictions are.

    - **Root Mean Squared Error (RMSE):** Similar to MAE but emphasizes larger errors, providing a more comprehensive view of prediction accuracy.

    - **R-squared (R2) Score:** It shows how much of the variance in the data is explained by the model, indicating how well the model fits the data.

2. **Classification Model Evaluation Metrics:** These metrics evaluate how accurately classification models classify data into different categories. They include:

    - **Accuracy:** This measures the overall correctness of the model's predictions.

    - **Precision:** It shows how many positive predictions were actually correct, focusing on the accuracy of positive predictions.

    - **Recall (Sensitivity):** This metric indicates how many actual positive instances the model correctly identified, emphasizing the model's ability to capture all positive cases.

    - **F1 Score:** The F1 score combines precision and recall into a single value, offering a balanced assessment of the model's performance in handling class imbalances.

- **Confusion Matrix:** A confusion matrix is a tabular representation of a machine learning model's performance, displaying the counts of true positive, true negative, false positive, and false negative predictions.

Explaining Each Evaluation Metric in Detail

1. Mean Absolute Error (MAE):

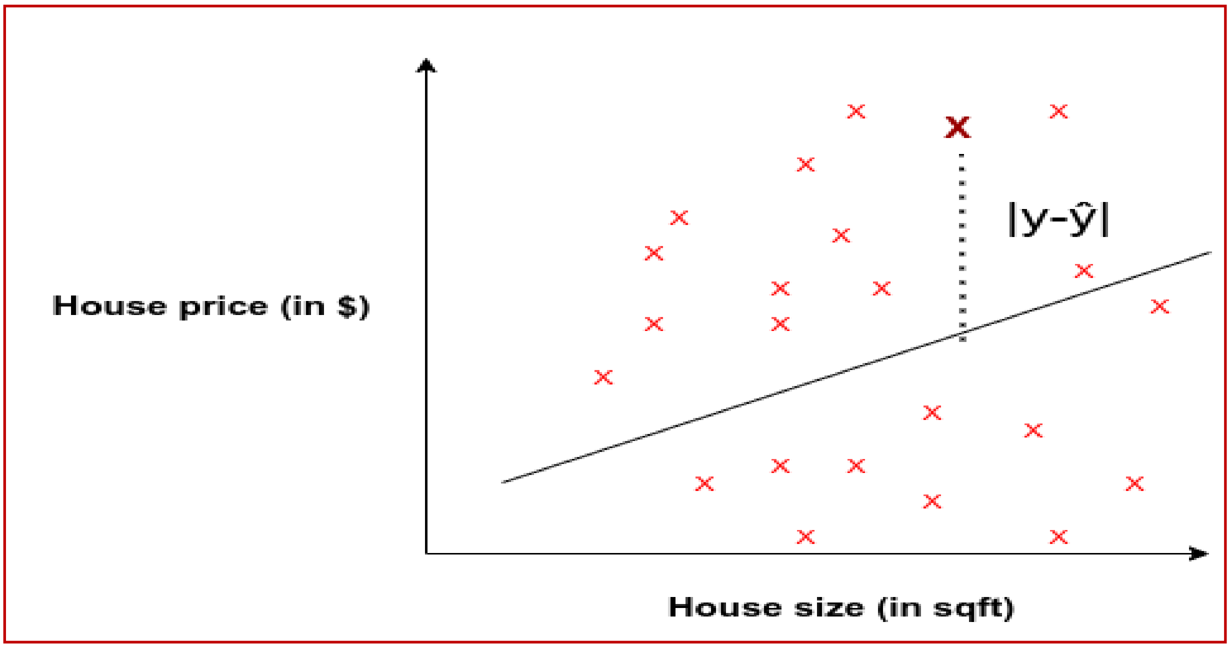
- MAE measures the average magnitude of errors between predicted and actual values in regression models.
- It provides insights into how accurate the predictions of the model are on average.
- MAE is calculated by taking the average of the absolute differences between predicted and actual values.

Formula for MAE:

$$\boxed{MAE = \frac{1}{N}\sum_{j=1}^{N}|y_j - \check{y}_j|}$$

Where:

- y_j: ground-truth value
- y_hat: predicted value from the regression model
- N: number of datums



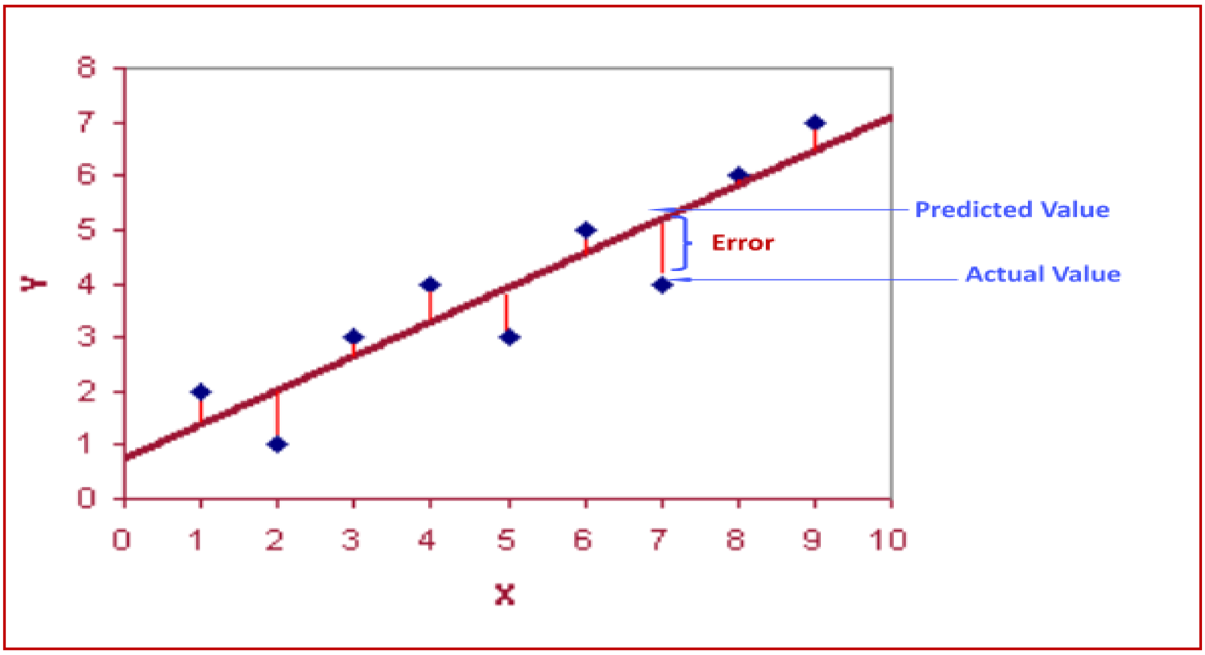
Example Graph: Mean Absolute Error

2. Root Mean Squared Error (RMSE):

- RMSE is similar to MAE but gives more weight to larger errors, making it sensitive to outliers.
- It provides a more comprehensive assessment of model performance by penalizing significant errors.
- RMSE is calculated by taking the square root of the average of squared differences between predicted and actual values.

Formula for RMSE:

$$RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y_j - \breve{y}_j)^2}$$



Example Graph: Root Mean Squared Error

3. **R-squared (R2) Score:**

- R2 score quantifies the proportion of variance in the dependent variable explained by independent variables in regression models.

- It indicates the goodness of fit of the regression model, showing how well the model fits the data.

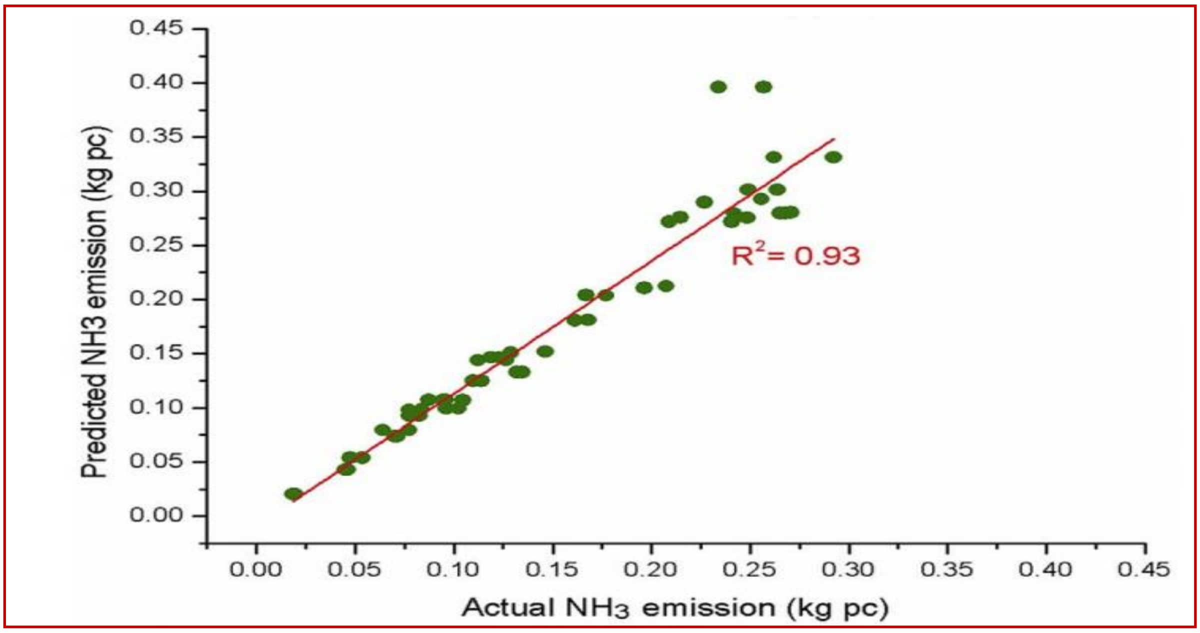
- R2 score ranges from 0 to 1, where 1 indicates a perfect fit and 0 indicates no relationship between variables.

Formula for R2 Score:

$$\begin{aligned} R^2 &= 1 - \frac{\text{sum squared regression (SSR)}}{\text{total sum of squares (SST)}}, \\ &= 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}. \end{aligned}$$

Where:

- $\bar{y}$ is the mean of the actual values.



Example Graph: R-Squared Score

4.Accuracy:

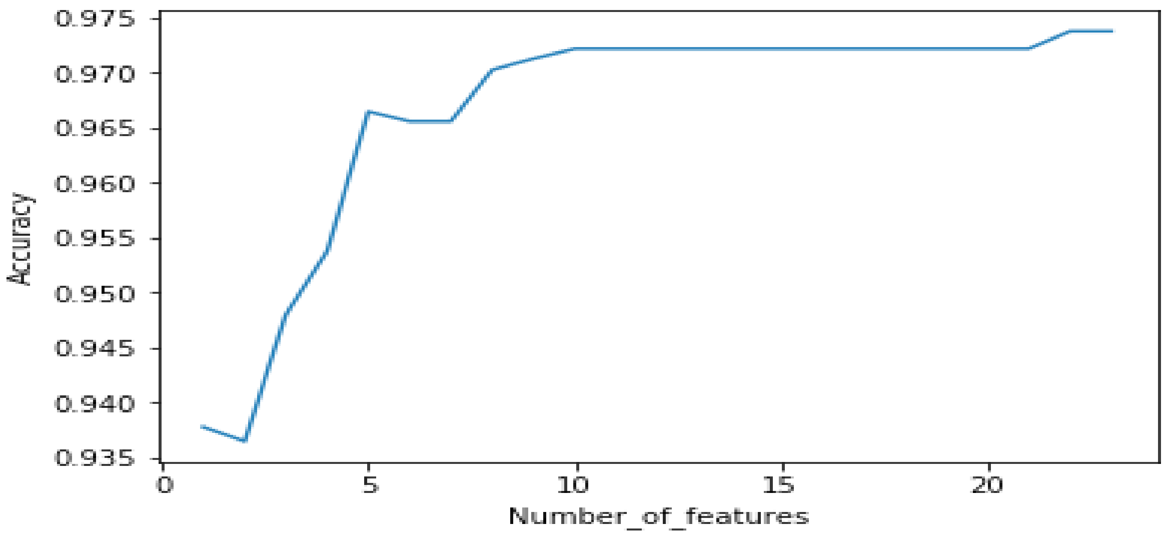
- Accuracy is a simple and intuitive metric that measures the percentage of correct predictions made by a model.
- It is suitable for balanced datasets where the positive and negative classes are similar in number.
- However, in imbalanced datasets, accuracy can be misleading as it favors the majority class predictions, neglecting the minority class.
- This can lead to an inaccurate assessment of the model's performance, especially in scenarios where the minority class is of high importance.

Formula for Accuracy:

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

Where:

- TP=True Positive
- TN=True Negative
- FP=False Positive
- FN=False Negative



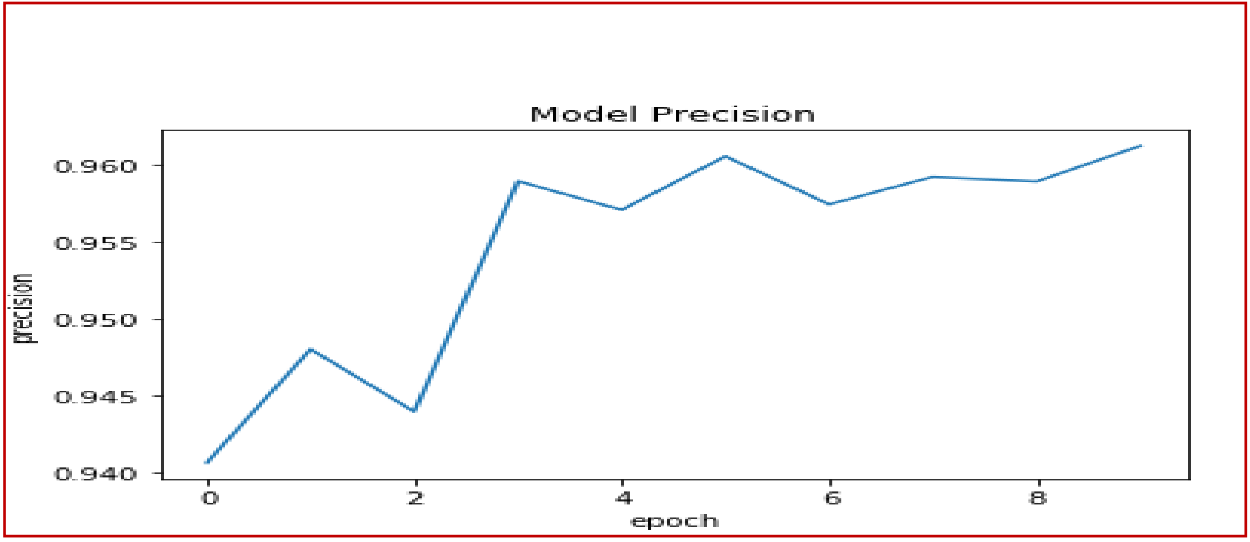
Example Graph: Accuracy

5.Precision:

- Precision measures the proportion of correctly predicted positive instances out of all instances predicted as positive.
- It is particularly valuable when the cost of false positives is significant.
- For instance, in medical diagnosis, high precision indicates accurate identification of patients with a disease, reducing false positive cases.

Formula for Precision:

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$



Example Graph: Precision

6.Recall:

- Recall (also known as Sensitivity) measures the proportion of correctly predicted positive instances out of all actual positive instances.
- It is crucial in scenarios where capturing all positive cases is vital, even if it results in some false alarms.
- For instance, in healthcare, high recall ensures that the model doesn't miss identifying patients with a disease, even if it means some healthy individuals are flagged for further evaluation.

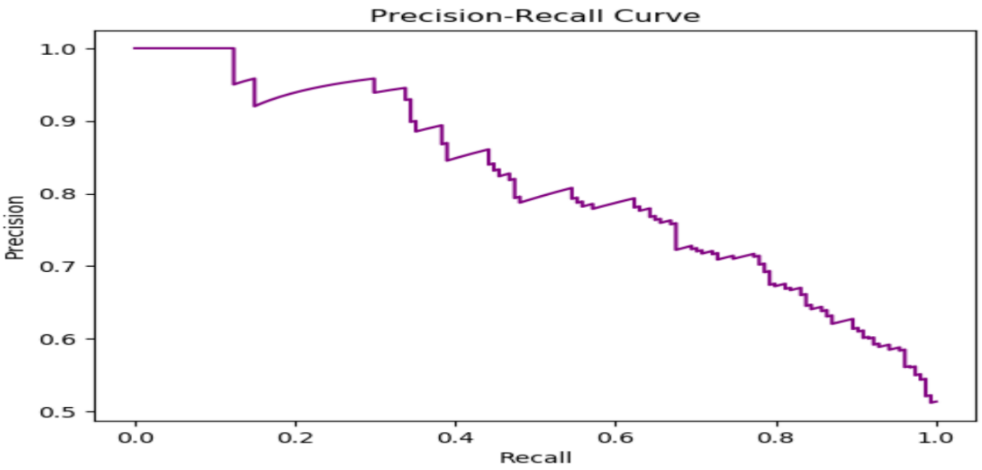
Formula for Recall:

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

Where:

TP=True Positive

FN=False Negative

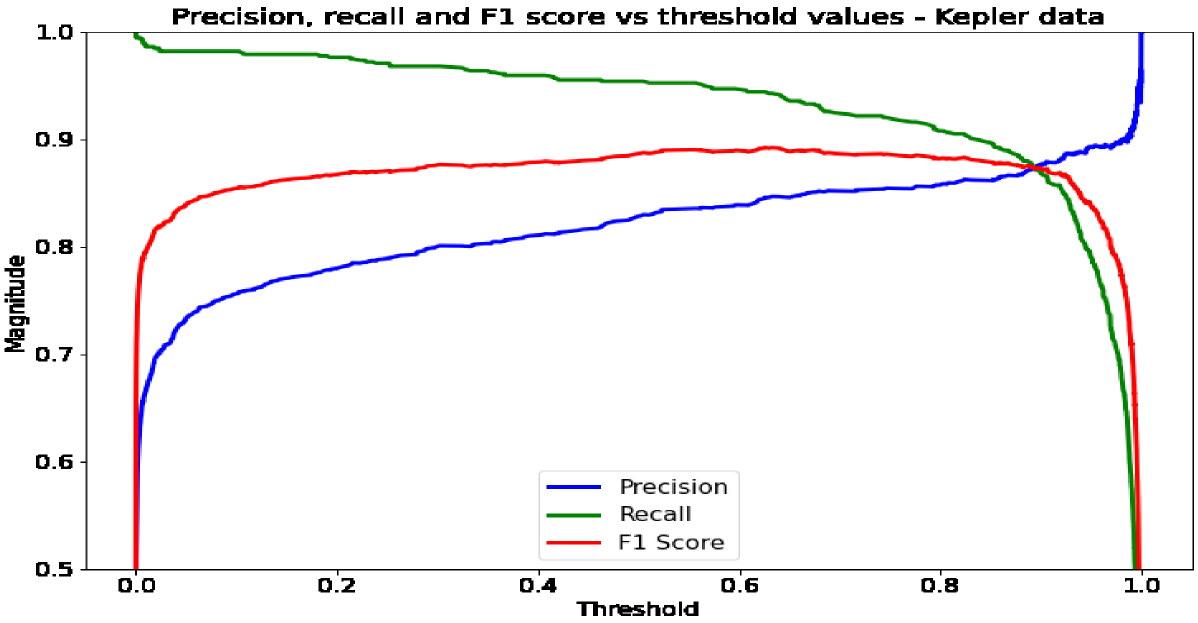


Example Graph: Recall

7.F1 Score:

- The F1 score represents the harmonic mean of precision and recall.
- It offers a balanced assessment of a model's performance by taking into account both precision and recall.
- Models with a similar balance between precision and recall are favored by the F1 score.
- The harmonic mean is particularly suitable for averaging ratios of values, making the F1 score valuable in scenarios with imbalanced precision and recall values.

$$F1\ score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = 2 \cdot \frac{Precision * Recall}{Precision + Recall}$$

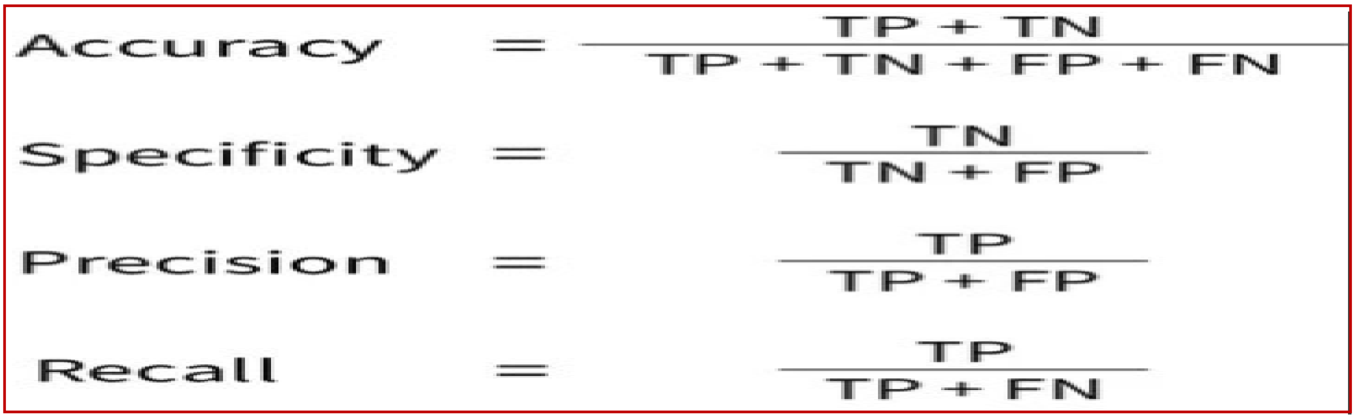


Example Graph: F1 Score

8.Confusion Matrix:

- The confusion matrix is a tabular representation of true and predicted classes in a classification problem.
- It displays the four possible combinations of true positives, true negatives, false positives, and false negatives, offering insights into the model's performance and areas for improvement.

Formula for Confusion Matrix:

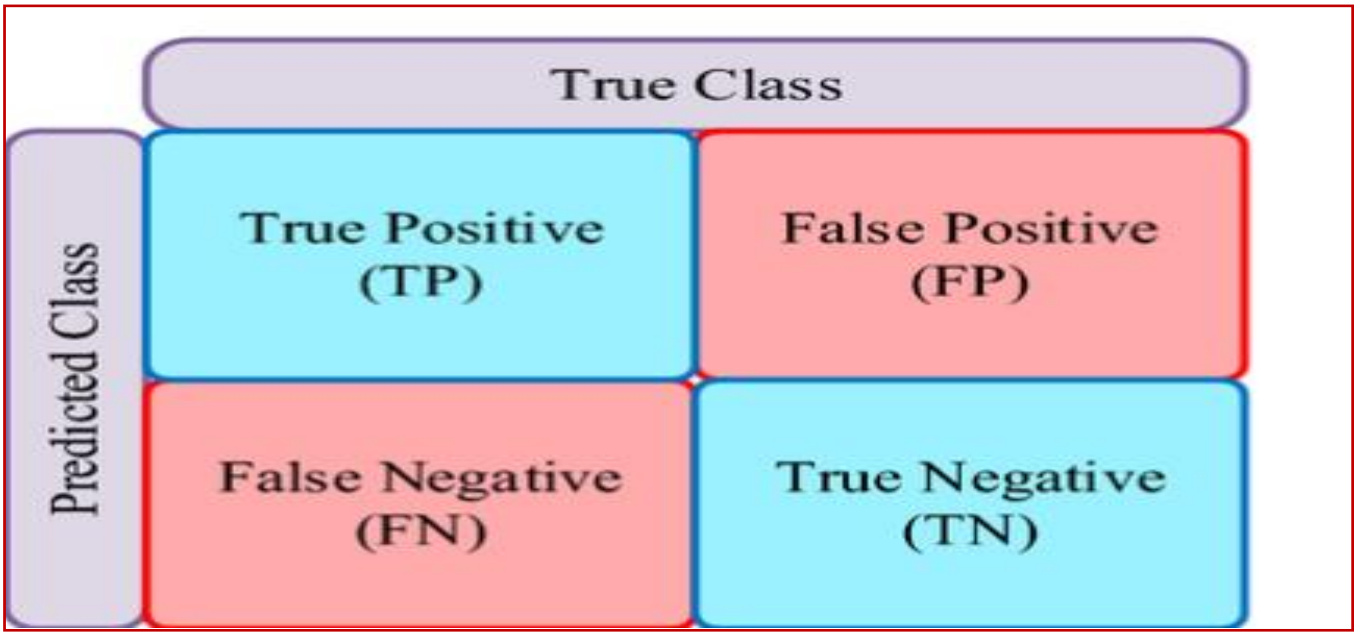


$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

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

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

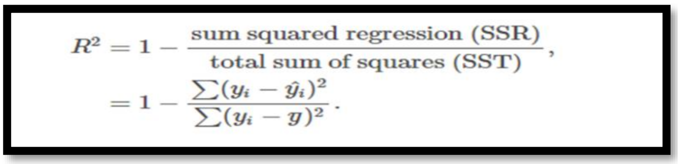
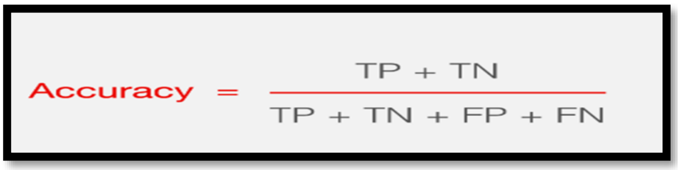
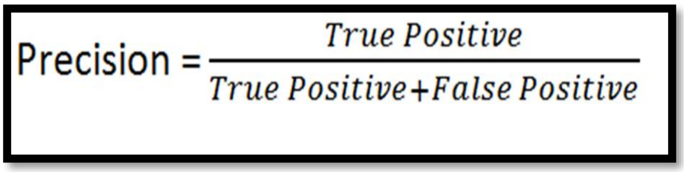
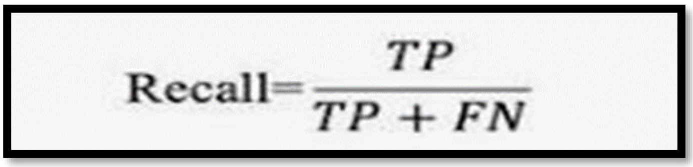
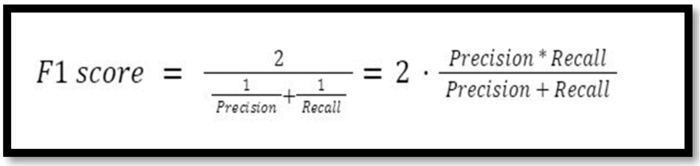
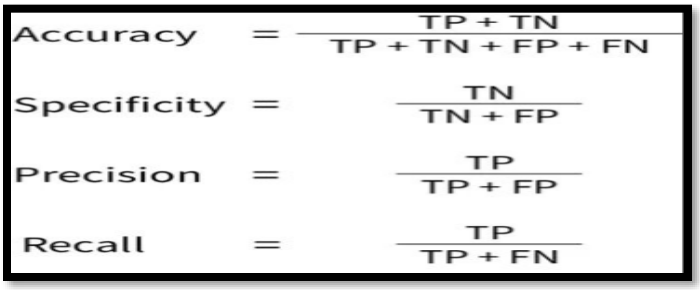


Confusion Matrix

Recap:

In conclusion, evaluation metrics are indispensable tools in ML for assessing model performance and guiding decision-making processes. Incorporating both regression and classification model evaluation metrics provides a comprehensive understanding of a model's capabilities and areas for improvement. The table below summarizes the key evaluation metrics discussed in this article, along with their descriptions and formulas:

Metric	**Description**	**Formula**
Mean Absolute Error	**Measures average magnitude of errors between predicted and actual values in regression models.**	$\boxed{MAE = \frac{1}{N}\sum_{j=1}^{N}\lvert y_j - \breve{y}_j\rvert}$
Root Mean Squared Error	**Similar to MAE but penalizes large errors more, sensitive to outliers.**	$\boxed{RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y_j - \breve{y}_j)^2}}$

R-squared (R2) Score	**Quantifies proportion of variance in dependent variable explained by independent variables.**	 $R^2 = 1 - \frac{\text{sum squared regression (SSR)}}{\text{total sum of squares (SST)}},$ $= 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}.$
Accuracy	**Measures proportion of correct predictions made by model over total predictions.**	 $\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$
Precision	**Measures proportion of correctly predicted positive instances out of all predicted positives.**	 $\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$
Recall (Sensitivity)	**Measures proportion of correctly predicted positive instances out of all actual positives.**	 $\text{Recall} = \frac{TP}{TP + FN}$
F1 Score	**Combines precision and recall into single value, offering balanced model performance assessment.**	 $F1\ score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = 2 \cdot \frac{Precision * Recall}{Precision + Recall}$
Confusion Matrix	**Tabular representation of true and predicted classes, displaying counts of true positives, true negatives, false positives, and false negatives.**	 $\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$ $\text{Specificity} = \frac{TN}{TN + FP}$ $\text{Precision} = \frac{TP}{TP + FP}$ $\text{Recall} = \frac{TP}{TP + FN}$

By leveraging evaluation metrics effectively, data scientists and ML practitioners can develop robust and accurate ML models that meet the desired performance standards and deliver reliable results in real-world applications.

Thank you.