

SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore - 641 107 Accredited by NAAC-UGC with 'A' Grade Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of Information Technology

Course Name – Data Warehouse & Mining

II Year / IV Semester

Topic – Prediction





Prediction

"prediction accuracy" refers to the degree to which a model correctly predicts outcomes on new data.

Typically measured using metrics like overall accuracy, precision, recall, F1-score, or more specialized metrics depending on the problem domain.

while "error measures" represent the opposite, quantifying the degree of incorrect predictions made by the model, often calculated as the difference between predicted and actual values.





Precision

It refers to the proportion of correct positive predictions (True Positives) out of all the positive predictions made by the model (True Positives + False Positives). It is a measure of the accuracy of the positive predictions.

The formula for Precision is:

$$Precision = \frac{Tri}{TruePositin}$$

A high Precision means that the model makes few False Positives. This metric is especially useful when the cost of false positives is high such as email spam detection.



uePositives ves+FalsePositives



Recall

It is also known as Sensitivity or True Positive Rate where we measures the proportion of actual positive instances that were correctly identified by the model. It is the ratio of True Positives to the total actual positives (True Positives + False Negatives).

The formula for Recall is:

 $Recall = \frac{TruePositives}{TruePositives+FalseNegatives}$

A high Recall means the model correctly identifies most of the positive instances, which is critical when False Negatives are costly, like in medical diagnoses.





Confusion Matrix

To better understand Precision and Recall which summarizes the performance of a classifier in four essential terms:

True Positives (TP): Correctly predicted positive instances.

False Positives (FP): Incorrectly predicted positive instances.

True Negatives (TN): Correctly predicted negative instances.

False Negatives (FN): Incorrectly predicted negative instances.





Mean Squared Error (MSE):

The average of the squared differences between predicted and actual values, commonly used for regression tasks.

Root Mean Squared Error (RMSE):

The square root of MSE, providing a more interpretable error measure in the same units as the target variable.

Mean Absolute Error (MAE):

The average absolute difference between predicted and actual values, less sensitive to outliers than MSE.







Evaluating Ensemble Models

Diversity:

A key aspect of ensemble methods, where individual models within the ensemble are encouraged to make different mistakes, leading to improved overall accuracy.

Bagging:

Creating multiple models by randomly sampling data with replacement from the original dataset, then averaging their predictions.

Boosting:

Sequentially building models where each new model focuses on correcting errors made by previous models, often using a weighting scheme.

Stacking:

Combining predictions from different base models by training a meta-model on their outputs.





THANK YOU

