

# Glossary

## Module 2 Lesson 2: From Modeling to Evaluation



Welcome! This alphabetized glossary contains many of the terms you'll find within this lesson. These terms are important for you to recognize when working in the industry, when participating in user groups, and when participating in other certificate programs.

Term	Definition
Binary classification model	A model that classifies data into two categories, such as yes/no or stop/go outcomes.
Data compilation	The process of gathering and organizing data required for modeling.
Data modeling	The stage in the data science methodology where data scientists develop models, either descriptive or predictive, to answer specific questions.
Descriptive model	A type of model that examines relationships between variables and makes inferences based on observed patterns.
Diagnostic measure based tuning	The process of fine-tuning the model by adjusting parameters based on diagnostic measures and performance indicators.
Diagnostic measures	The evaluation of a model's performance of a model to ensure that the model functions as intended.
Discrimination criterion	A measure used to evaluate the performance of the model in classifying different outcomes.
False-positive rate	The rate at which the model incorrectly identifies negative outcomes as positive.
Histogram	A graphical representation of the distribution of a dataset, where the data is divided into intervals or bins, and the height of each bar represents the frequency or count of data points falling within that interval.
Maximum separation	The point where the ROC curve provides the best discrimination between true-positive and false-positive rates, indicating the most effective model.
Model evaluation	The process of assessing the quality and relevance of the model before deployment.
Optimal model	The model that provides the maximum separation between the ROC curve and the baseline, indicating higher accuracy and effectiveness.
Receiver Operating Characteristic (ROC)	Originally developed for military radar, the military used this statistical curve to assess the performance of binary classification models.
Relative misclassification cost	This measurement is a parameter in model building used to tune the trade-off between true-positive and false-positive rates.
ROC curve (Receiver Operating Characteristic curve)	A diagnostic tool used to determine the optimal classification model's performance.
Separation	Separation is the degree of discrimination achieved by the model in correctly classifying outcomes.
Statistical significance testing	Evaluation technique to verify that data is appropriately handled and interpreted within the model.
True-positive rate	The rate at which the model correctly identifies positive outcomes.

### Author(s)

[Dr. Pooja](#)  
[Patsy Kravitz](#)

### Changelog

Date	Version	Changed by	Change Description
2023-08-03	0.1	Patsy Kravitz	Initial version created

© IBM Corporation 2023. All rights reserved.