PriMera Scientific Engineering
Volume 3 Issue 6 December 2023
ISSN: 2834-2550

PriMera Scientific Publications

Measurement of a Classification Model

Type: Editorial
Received: November 18, 2023
Published: November 23, 2023

Citation:
Harco Leslie Hendric Spits Warnars. "Measurement of a Classification Model". PriMera Scientific Engineering 3.6 (2023): 01-02.

Harco Leslie Hendric Spits Warnars*

Prof. in Computing, Computer Science Department, BINUS Graduate Program - Doctor of Computer Science, Bina Nusantara University, Jakarta, Indonesia

***Corresponding Author:** Harco Leslie Hendric Spits Warnars, Prof. in Computing, Computer Science Department, BINUS Graduate Program - Doctor of Computer Science, Bina Nusantara University, Jakarta, Indonesia.

When dealing with supervised models containing classification and regression, we need special matrices, such as confusion matrices. In software engineering, the final step, such as testing, is the most crucial stage in software creation, and that is the confusion matrix. We must have software that has been tested, and in this case, it is called white box testing, which is testing carried out from the internal side of the software. It is essential that software is not only created but can also be tested internally and even better if it is also tested externally, which is also often referred to as black box testing. Black box testing is a software testing process carried out by distributing questionnaires to software users as a user acceptance test.

A well-known classification algorithm we often hear about and use in solving several problems is the k-nearest Neighbor (KNN) Classifier algorithm, which can also be used in regression processes such as the KNN regression algorithm. Also, the Naïve Bayes Classifier algorithm, including the Artificial Neural Network (ANN) algorithm, where the Neural Network (NN) algorithm can also carry out the regression process. Furthermore, the Support Vector Machine (SVM) algorithm can also conduct the regression process with the SVM regression algorithm and the Decision Tree Classifier (DT) algorithm, including the regression process with the DT regression algorithm. Apart from that, the random forest classifier algorithm can also be used for classification, and the regression process is carried out with the RF regression algorithm. Furthermore, the classification process can also be carried out using the Generalized Regression NN (GRNN) algorithm as a variation of the Radial Basis NN (RBNN) algorithm, where GRNN can be used for classification or regression. Finally, the classification process can also use the algorithm. The gradient-boosted tree classifier can also be chosen to carry out the classification. The regression process can also be done using the Gradient gradient-boosting machine-regression algorithm. Furthermore, the classification process can be carried out with multilayer perceptron Classifier algorithms, One-vs-Rest (aka One-vs-All) Classifier algorithms, and factorisation machine Classifier algorithms.

In addition, the classification process can be used to classify with two output results, namely the binary classification algorithm. The following classification algorithms can be used for binary classification and regression processes: the Logistic Regression algorithm, LightGBM algorithm, XGBoost algorithm and Neural Network (NN) algorithm (Deep Learning). Apart from that, the following binary classification algorithms are only specific for classifying with two output results. They are the Naive

Bayes (Gaussian) algorithm, the Naive Bayes (Bernoulli) algorithm, the K Nearest Neighbors (KNN) algorithm, the Support Vector Machine (SVM) algorithm, the Decision Tree (DT) algorithm, Random Forest (RF) algorithm and Gradient Boosting Machine algorithm.

A classification model can be measured using a confusion matrix, a Matrix for Summarizing the performance of the Classification algorithm and is known as an error matrix. A confusion matrix is a matrix between the predicted and actual conditions in the population. The predicted condition has PP as the number of positively predicted cases in the population, whilst PN is the number of negatively predicted cases in the population. Meanwhile, the actual condition has P as the number of positive actual cases in the population and N as the number of negative actual cases. The confusion matrix has 4 scores: TP as True Positive, TN as True Negative, FN as False Negative and FP as False Positive. TP is a test result that correctly indicates the presence of a condition or characteristic, whilst TN is a test result that accurately displays the absence of a condition or characteristic. Meanwhile, FN is a test result that wrongly indicates that a particular condition or attribute is absent, and FP is a result that wrongly suggests that a specific condition or attribute is present.

The following are metrics that can be used in measuring classification models, and they are:

1. Accuracy.
2. Classification Error Rate.
3. True Positive Rate (TPR).
4. False Positive Rate (FPR).
5. False Negative Rate (FNR).
6. True Negative Rate (TNR).
7. Positive Predictive Value (PPV).
8. Negative Predictive Value (NPV).
9. False Discovery Rate (FDR).
10. False Omission Rate (FOR).
11. Misclassification Rate.
12. F1 Score.
13. Kappa Cohen.
14. Matthew Correlation Coefficient (MCC).
15. +LR (Positive Likelihood Ratio).
16. -LR (Negative Likelihood Ratio).
17. J Youden.
18. G Measure.
19. D2H (Distance to Heaven).
20. AUC=Area under the Receiver Operating Characteristic (ROC) curve.

This metric can only be applied to the classification process. It cannot possibly be used for other supervised methods, such as regression, and is even less applied to clustering processes as unsupervised models.