

Evaluation metrics for classification models

Why do we need evaluation metrics for classification models?

- In machine learning on broad aspect, we have a problem statement to address, then we fetch data related to it, perform analysis and feature engineering. Then use the data to train the model which we finally use to do the prediction.
- Thus, the output of trained machine learning model is consumed by end users for making their decisions.
- In our cybersecurity use case, the impact of the models becomes extremely critical because of the nature of outcome helps to make important decisions about benign and malicious events or type of malicious events.
- In order to use machine learning models in real world scenario, we need to address the fundamental questions such as: -
 1. Why should the end user trust the trained model?
 2. How does our model perform relative to the other models trained by others?
- To address the above fundamental questions, we need to define the governance and framework of evaluation of models which help us understand the given model's performance and also compare them on reliable and useful metrics with other models, which finally allows the end users to make decisions on determining the quality of output produced by the given model and describe the same in detail.
- In terms of building structure for evaluation of classification models, we need to perform seven major tasks: -
 1. List the metrics that can be used for the use case.
 2. Define each metric in detail and explain its benefits and limitations (if any).
 3. Document the evaluation results of all previous models observed from literature survey.
 4. Compute the performance of our model based on each metric defined in task 2.
 5. Quantitatively document the comparison of performance of our model with previously trained models observed in literature survey.
 6. Describe the performance of our model with respect to previously trained model using the data documented in task 6. We need to compute the gap between performance of our model with respect to other models for all available metrics.
 7. Derive the inferences based on task 5 and task 6, explain reason for the same. If our model performs better than previously trained models, we need to explain the reasons for achieving better results. Similarly, if our model performs worse than previously trained models, we need to identify the gaps that we need to work on to reach that performance.
- Robust documentation of the above tasks will help us define the performance of our models and also help end users to decide about usage of the models.

List of metrics for evaluation of classification models (both binary and multi-class)

1. Confusion Matrix
2. Accuracy
3. Precision
4. Recall
5. F1-Score
6. ROC curve
7. AUC score
8. Specificity
9. Balanced accuracy
10. Matthews Correlation Coefficient (MCC)
11. Logarithmic Loss
12. Binary Crossentropy
13. Categorical Crossentropy
14. Concordance and Discordance
15. Somers-D Statistic
16. Gini coefficient
17. Type 1 error
18. Type 2 error
19. Negative predictive value
20. False discovery rate
21. Cohen kappa metric
22. Precision – Recall curve
23. Brier score

Definition and details about each metric: -

Evaluation results as per literature survey: -

Documentation of performance of models based on each metric: -

Quantitative comparison of performance of our models with respect to models observed in literature survey: -

Descriptive comparison of performance of our models with respect to models observed in literature survey: -

Inference about the performance of our model: -

Sources for evaluation metrics: -

<https://www.kdnuggets.com/2020/04/performance-evaluation-metrics-classification.html>

<https://www.explorium.ai/blog/machine-learning/top-10-evaluation-metrics-for-classification-models/>

<https://www.appsilon.com/post/machine-learning-evaluation-metrics-classification>

<https://community.alteryx.com/t5/Data-Science/Metric-Matters-Part-1-Evaluating-Classification-Models/ba-p/719190>

<https://www.deepchecks.com/a-guide-to-evaluation-metrics-for-classification-models/>

<https://www.evidentlyai.com/classification-metrics/multi-class-metrics>

<https://cloud.google.com/vertex-ai/docs/video-data/classification/evaluate-model>

https://www.mlwhiz.com/p/eval_metrics

<https://www.datacamp.com/blog/classification-machine-learning>

<https://www.analyticsvidhya.com/blog/2021/07/metrics-to-evaluate-your-classification-model-to-take-the-right-decisions/>

<https://www.machinelearningplus.com/machine-learning/evaluation-metrics-classification-models-r/>

<https://neptune.ai/blog/evaluation-metrics-binary-classification>

<https://www.kdnuggets.com/2020/05/model-evaluation-metrics-machine-learning.html>

https://cs229.stanford.edu/lectures-spring2022/cs229-evaluation_metrics_slides.pdf

<https://www.kdnuggets.com/understanding-classification-metrics-your-guide-to-assessing-model-accuracy>

<https://www.appliedaicourse.com/blog/evaluation-metrics-in-machine-learning/>

https://colab.research.google.com/github/datacommonsorg/api-python/blob/master/notebooks/intro_data_science/Classification_and_Model_Evaluation.ipynb

<https://keylabs.ai/blog/overview-of-evaluation-metrics-for-classification-models/>

<https://thesai.org/Publications/ViewPaper?Volume=12&Issue=6&Code=ijacsa&SerialNo=70>

<https://deepai.org/machine-learning-glossary-and-terms/evaluation-metrics>

<https://www.youtube.com/watch?v=5vqk6HnITko&list=PLiteiKUvOPTelAliquq5-VKFguUJpv1BT>

<https://www.youtube.com/watch?v=LbX4X71-TFI>

https://www.youtube.com/watch?v=lt1YxJ_8Jzs

https://www.youtube.com/watch?v=7ZfcoEao_5c

<https://www.youtube.com/watch?v=PeYQlyOyKB8>

<https://www.youtube.com/watch?v=BEeXPgfhCsA>

<https://www.youtube.com/watch?v=wpQiEHYkBys>

<https://www.youtube.com/watch?v=sgv1Q46Tdmw>