MachineLearnAthon - Microlecture Evaluation metrics

Recorded by Lara Kuhlmann

MachineLearnAthon
A project Co-funded by the Erasmus+ programme of the European Union

Learning outcomes of today

After successfully completing this micro-lecture, you are able to….

- Explain the significance of evaluation metrics in assessing model performance
- Assess the suitability of specific metrics for different types of machine learning problems (e.g., classification vs. regression)
- Interpret classification metrics such as Accuracy, Precision, and Recall using confusion matrix values
- Calculate classification performance using TP, FP, TN, and FN values
- Apply regression metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) to quantify model prediction errors

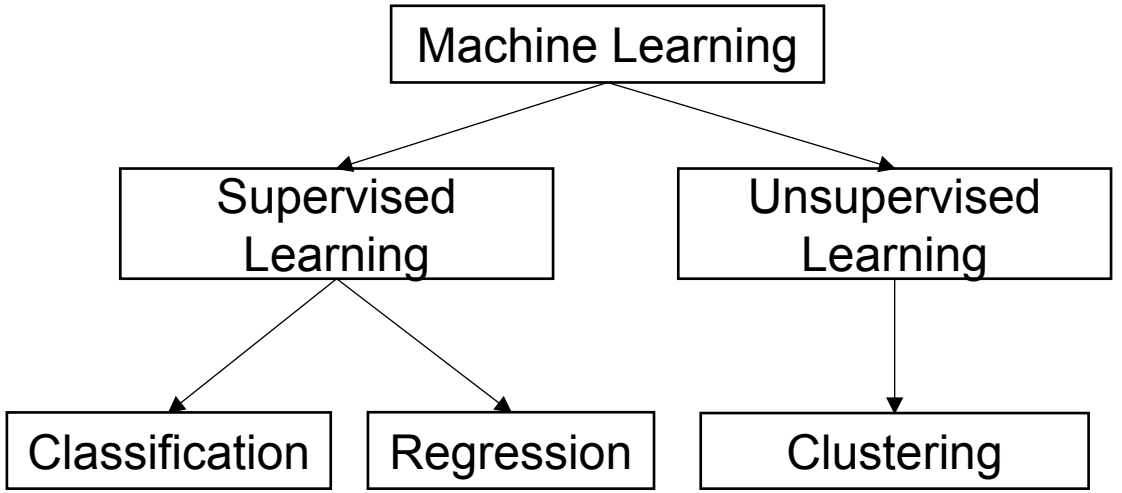
Agenda for today

- Evaluation metrics in Machine learning
- Classification Metrics
  - Accuracy
  - Precision, Recall
- Regression Metrics
  - Mean Absolute Error (MAE)
  - Mean Squared Error (MSE)
  - Root Mean Squared Error (RMSE)

technische universität dortmund
TU Delft
Co-funded by the Erasmus+ programme of the European Union
Machine LearnAthon

Evaluation metrics in Machine learning

- Evaluation metrics are essential in machine learning for assessing model performance. These metrics help determine how well a model is likely to perform in real-world application.
- Different tasks like classification, regression, and clustering each have unique metrics tailored to their requirements and objectives.
- Classification metrics typically evaluate accuracy and balance; regression metrics focus on error magnitude, while clustering metrics assess grouping quality.
- Evaluation metrics should be applied on a test set, which was not used for training the machine learning model.



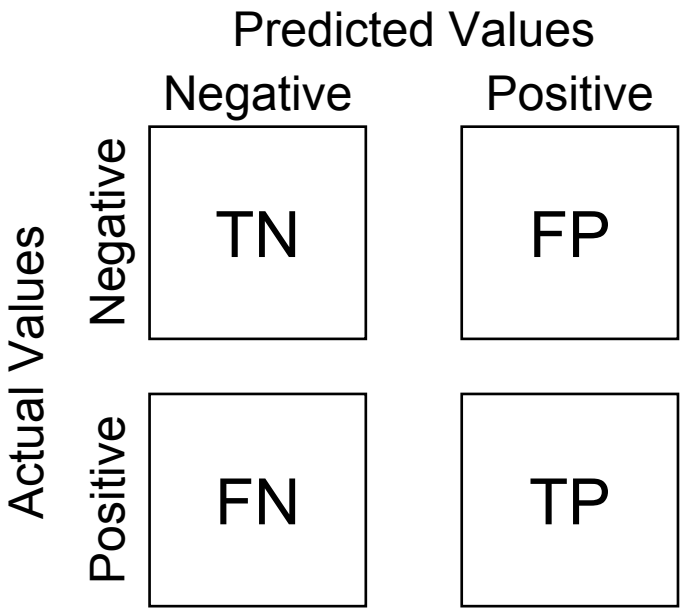
Zheng, A. (2015)

Co-funded by the Erasmus+ programme of the European Union

Classification Metrics - Accuracy

Accuracy measures the proportion of correct predictions over total predictions, often used in balanced datasets for a clear view of correct classifications.

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



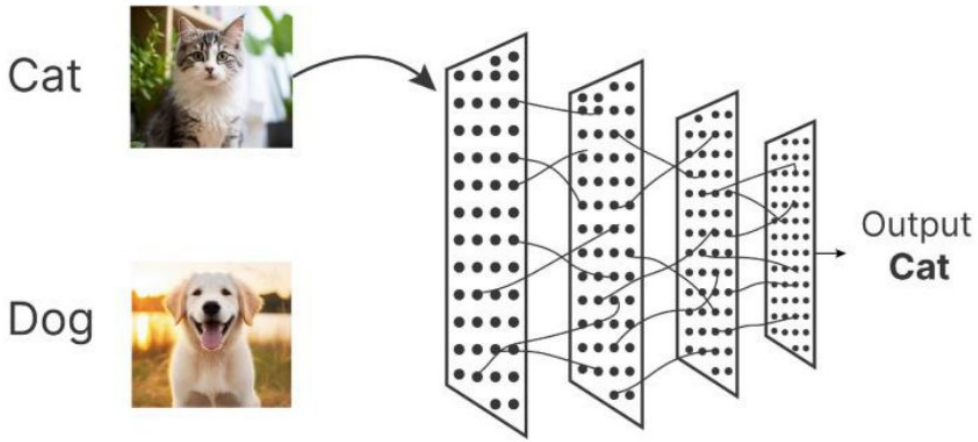
Zheng, A. (2015)

technische universität dortmund
TU Delft
Co-funded by the Erasmus+ programme of the European Union
Machine LearnAthon

Classification Metrics - Accuracy

In an example, a model examines 100 images to classify dogs and cats: 59 are dogs (48 correctly identified, 11 misclassified) and 41 are cats (37 correctly identified, 4 misclassified)

$$\text{Accuracy} = \frac{48+37}{37+9+6+48} = 0.85 = 85\%$$



Actual Values \ Predicted Values	Cat	Dog
Cat	37	4
Dog	11	48

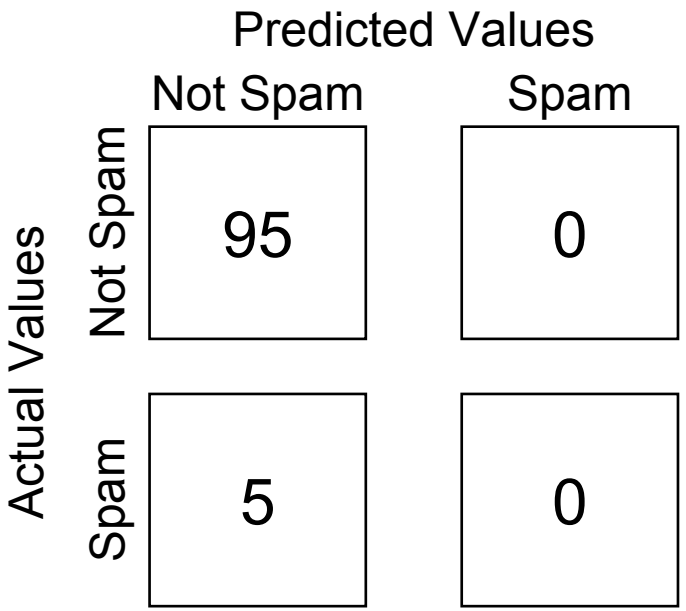
Naidu et al (2023), Zheng, A. (2015)

technische universität dortmund
TU Delft
Co-funded by the Erasmus+ programme of the European Union
Machine LearnAthon

Classification Metrics - Accuracy

- In another example, our model predicts whether an email is spam or not. Out of 100 emails, 95 are 'Not Spam' and 5 are 'Spam' in the dataset.

$$\text{Accuracy} = \frac{95+0}{0+5+95+0} = 0.95 = 95\%$$



- In imbalanced datasets, accuracy can mislead; alternative metrics like precision and recall may provide a better insight

Rainio et al (2024)

Co-funded by the Erasmus+ programme of the European Union

Classification Metrics - Precision, Recall

- Precision measures the proportion of true positive predictions out of all the positive predictions made by the model.

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

- Back to the example for predicting spam and not spam Emails:

$$\text{Accuracy} = \frac{90+3}{90+3+5+2} = 0.93 = 93\%$$

$$\text{Precision} = \frac{3}{2+3} = 0.6 = 60\%$$

		Predicted Values: Negative	Predicted Values: Positive
Actual Values	Negative	TN = 90	FP = 2
Actual Values	Positive	FN = 5	TP = 3

Naidu et al (2023), Rainio et al (2024)

Co-funded by the Erasmus+ programme of the European Union

Classification Metrics - Precision, Recall

- Recall, also known as sensitivity is the proportion of true positives out of all actual positive cases.

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

- Back to the example for predicting spam and not spam Emails:

$$\text{Accuracy} = 93\%$$

$$\text{Precision} = 60\%$$

$$\text{Recall} = \frac{3}{3+5} = 0.375 = 37.5\%$$

- Recall is extremely important in such cases that missing a positive instance is extremely costly. E.g. Fraud Detection

Actual Values \ Predicted Values	Negative	Positive
Negative	TN = 90	FP = 2
Positive	FN = 5	TP = 3

Naidu et al (2023), Rainio et al (2024)

Co-funded by the Erasmus+ programme of the European Union

Regression Metrics

- Difference to classification is a numeric target value, therefore other evaluation metrics are needed to measure the model performance
- Variety of metrics with different approaches and therefore other advantageous use cases
- Here: Mean Absolute Error, Mean Squared Error, Root Mean Squared Error

Tatachar (2021)

Co-funded by the Erasmus+ programme of the European Union

Regression Metrics - Mean Absolute Error (MAE)

- Mean Absolute Error (MAE) measures the average of the absolute difference between the observed and the predicted values

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left| Y_i - \hat{Y}i \right|$$

  - n: Number of predictions
  - Y_i : Observed values
  - $\hat{Y}i$: Predicted values

- Example: weather prediction

$$MAE = \frac{(|28 - 27| + |30 - 28| + |28 - 24| + |25 - 21|)}{4} = 2.75$$

	Actual temperature in °C	Predicted temperature in °C
1	28	27
2	30	28
3	28	24
4	25	21

Tatachar (2021)

Co-funded by the Erasmus+ programme of the European Union

Regression Metrics – Mean Squared Error (MSE)

- Mean Squared Error (MSE) measures the average squared difference between the observed and the predicted values

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}i\right)2$$

  - n: Number of predictions
  - Y_i : Observed values
  - $\hat{Y}i$: Predicted values

- Example: weather prediction

$$MSE = \frac{((28-27)2+(30-28)2+(28-24)2+(25-21)2)}{4} = 9.25$$

	Actual temperature in °C	Predicted temperature in °C
1	28	27
2	30	28
3	28	24
4	25	21

Plevris et al (2022), Tatachar (2021)

Co-funded by the Erasmus+ programme of the European Union

Regression Metrics – Root Mean Squared Error (RMSE)

- Root Mean Squared Error (RMSE) is the square root of the MSE

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}i\right)2}$$

  - n: Number of predictions
  - Y_i : Observed values
  - $\hat{Y}i$: Predicted values

- Example: weather prediction

$$RMSE = \sqrt{MSE} = \sqrt{9.25} \approx 3.041$$

	Actual temperature in °C	Predicted temperature in °C
1	28	27
2	30	28
3	28	24
4	25	21

Plevris et al (2022), Tatachar (2021)

Co-funded by the Erasmus+ programme of the European Union

Recap this lecture

After successfully completing this lecture, you are able to….

- Explain the significance of evaluation metrics in assessing model performance
- Assess the suitability of specific metrics for different types of machine learning problems (e.g., classification vs. regression)
- Interpret classification metrics such as Accuracy, Precision, and Recall using confusion matrix values
- Calculate classification performance using TP, FP, TN, and FN values
- Apply regression metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) to quantify model prediction errors

The European Commission support for the production of this publication does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

This material was created with the assistance of an AI language model.

technische universität dortmund
TU Delft
Co-funded by the Erasmus+ programme of the European Union
Machine LearnAthon