Supervised approaches for policy research

Edgar Salas Gironés,
e.girones@tudelft.nl

June 29, 2025

Supervised approaches

In unsupervised approaches, we focused on techniques that do not account for a 'target variable', what do we want to predict?

A few examples:

- We want to perform topic modeling, and we have an intuition of what topics will appear in the texts.
- We would like to classify texts based on some variables (e.g. relevant, non-relevant).
- We want to extract information based on certain categories.

Supervised learning 101

- For using supervised approaches, we need to have labeled examples.
- The task (and its evaluation is based on how good our model learns to predict the right label for new data).
- The models wants to learn how to associate features with labels

General structure of a supervised approach

Independent of the method or technique (e.g. using a decision tree or a neural network), the dataset is divided into three sets:

- Training set → The model learns patterns using this set.
- Test set. → The model performance is evaluated using this set.
- Validation set. → Are you already sure that the model performs well? A final evaluation is used using this set.[1]

---

[1]Another option is cross-validation, reference here

Evaluation
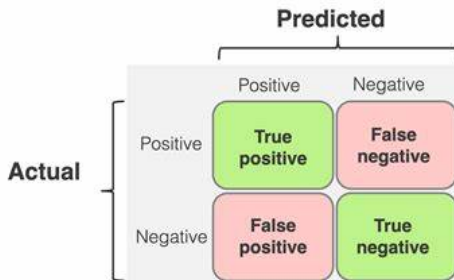
Four major evaluation metrics:

- Accuracy = How often was the model correct in prediction a label?
- Precision = Out of the predictions that model gave to a label (e.g. A), how many there were correct?
- Recall = Out of all entries with one label (e.g. A), how many where correctly identified?
- F1 score = 2 × (precision × recall) / (precision + recall)