

Challenges

- Data quality for learning
 - Effect of data collection processes (sampling, redundancies,...)
 - Effect of noise in the data/labels (e.g., distant supervision)
- Ethical/legal/risk considerations
 - How to evaluate ethical-/legal-/risk-aware models?
 - How to ensure statistical fairness, esp. for dynamic environments?

Methods

- “Data correction” methods
 - Changing the data distribution (massaging, reweighting, augmentation)
 - Local/global approaches
- In-design methods
 - Changing the learning algorithms
 - esp., for dynamic environments, so coupling with concept drifts