

Presentation Notes

Date: November 24, 2024

General Presentation framework:

- What is the topic, why does it matter for official statistics?
- Definitions and key methods explained simply
- examples
- contributions
- flowcharts, graphs, tables
- conclusion

1. Section 1: Estimation of the generalization error and its uncertainty

2. Section 2: Interpretable machine learning

(Same format as above) “Would you trust an ML model if you couldn’t explain its decision?”

3. Section 3: Machine learning for complex sample designs

(Same format as above)

4. Section 4: Quantitative methods for uncertainty quantification

(Same format as above)

5. Section 5: Machine learning operations and reproducibility in official statistics

(Same format as above)

6. Section 6: Fairness and Bias Auditing

Key Points:

- technical + social and societal challenges of ML in public sector
- what are the problems? ans: shifts in decision making responsibility and immense technical stability
- protected attributes
- measurement errors can also affect model training
- group, subgroup and individual fairness (concepts, and challenge)
- prediction and decision step (talking about how crucial the prediction step in official statistics is)
- The fairness of ADM systems starts way before decisions are made. Every step in the data pipeline - from designing surveys to cleaning and processing data - contributes to the fairness (or unfairness) of the final system
- Some effects may have bigger impact than others
- Two steps of ADM where prediction step directly affects the decision step
- integrating fairness aspects into existing quality criterion (contributions)
- the way ahead with the possible new opportunities (contribution)

Visual Ideas:

- ★ Example: Bias in a loan approval system
- ★ Data collection -> Cleaning -> Training -> Prediction -> Decision, with bias sources highlighted at each stage.
- ★ example of dealing with bias (the formula)
- ★ visually showing the way ahead?

Questions / topics to Address:

Q1: the connection between Interpretability and fairness of ML