

Statistical Bias in Machine Learning Applications

- In many real-world applications, ML is followed by aggregation:
 - *Predicting election outcomes based on social media data*
 - *Estimating area of ice sheets based on satellite images*
 - *Analysing spread of disease (epidemiology)*
- New insight: classification errors do not cancel out by aggregation
⇒ see Meertens et al., SDM19
- **Next step:** develop methods to further reduce statistical bias
- Both theoretical and data-driven approach is possible
 - *Possibility of internship at Statistics Netherlands (CBS, The Hague)*
 - *Supervised by Quinten Meertens (LIACS and CBS) and Frank Takes*