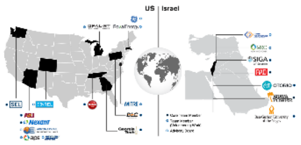
Task 12 - Explainable cyber A.I. analytics

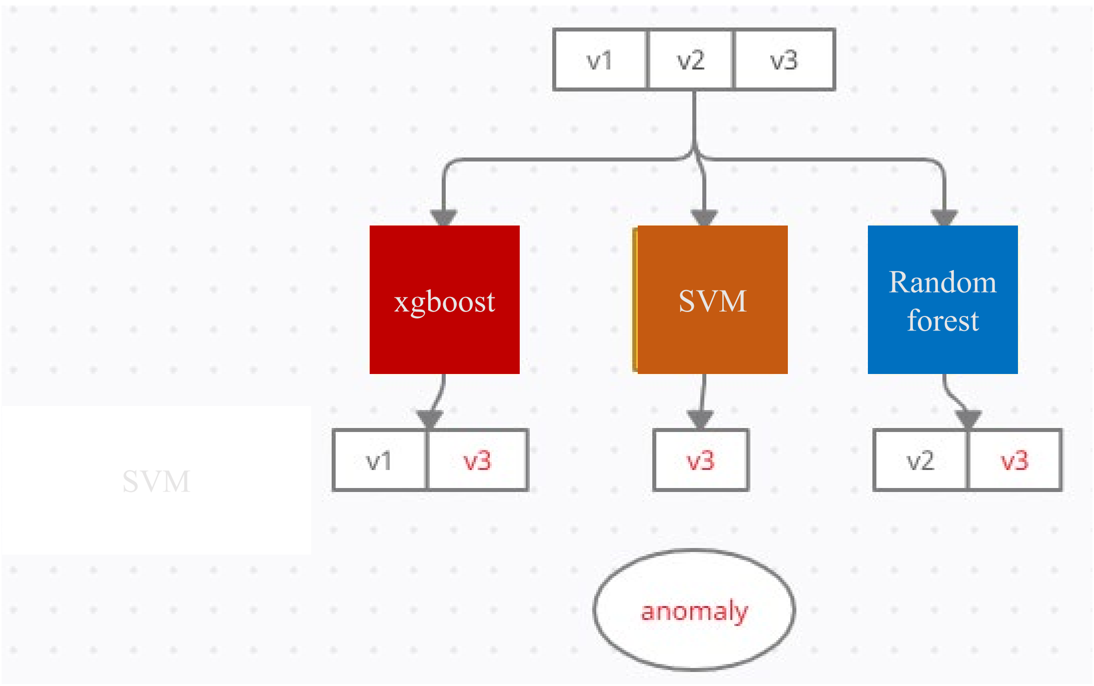
20/3/23

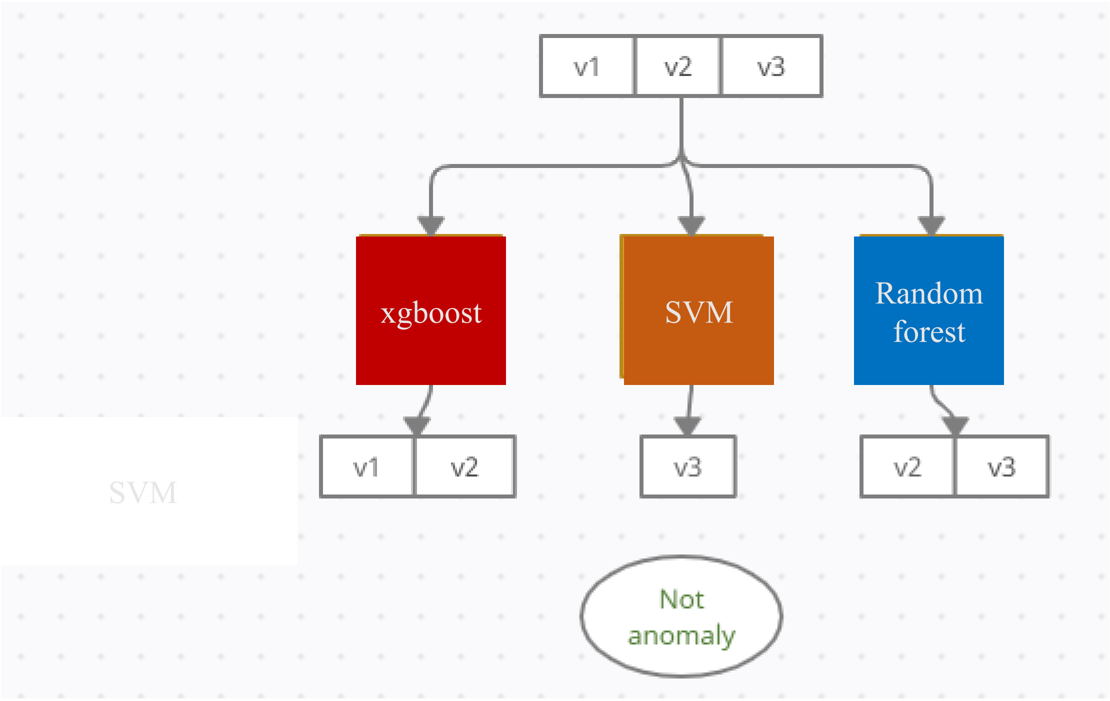
Ben Gurion University

Task 12 - Explainable cyber AI analytics

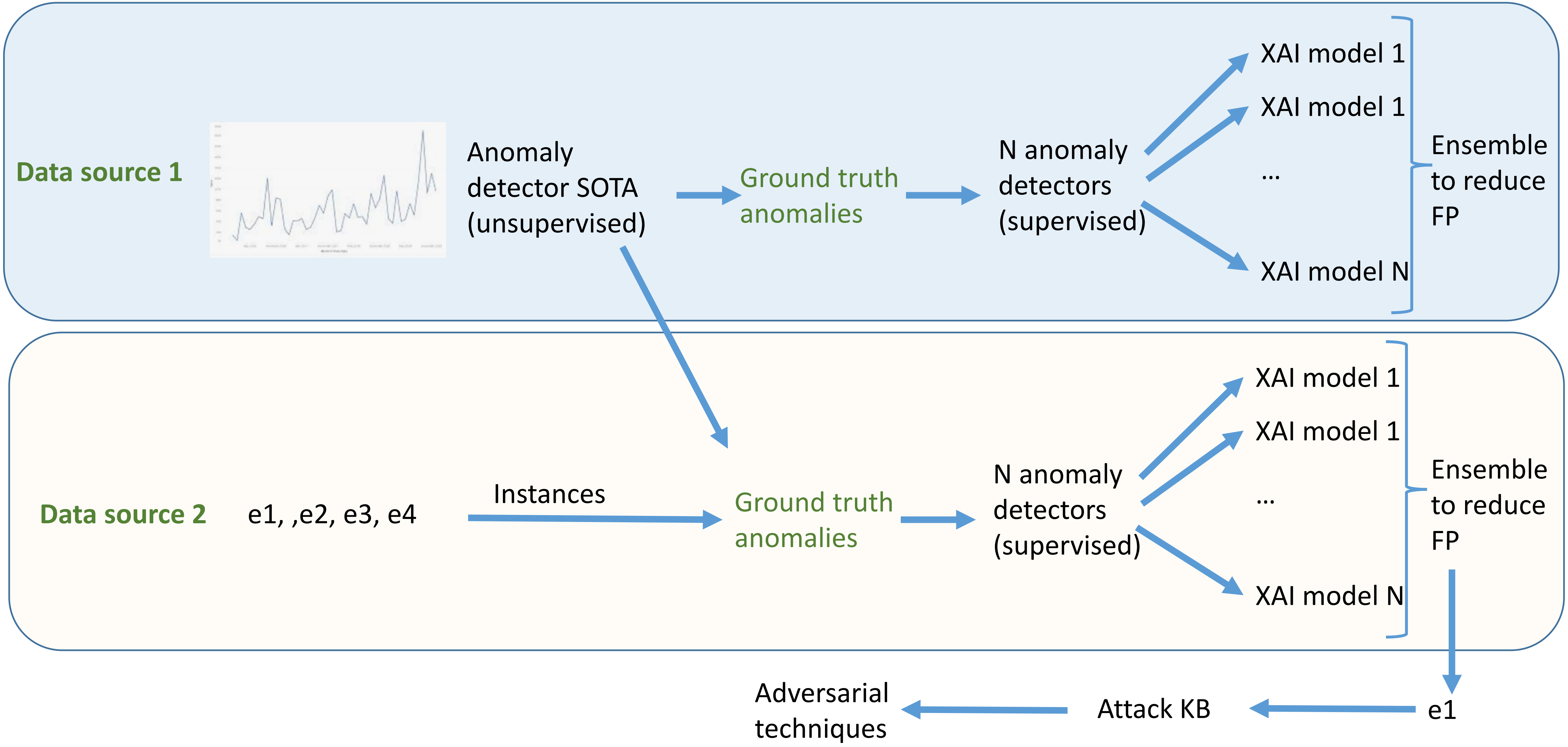


- **Method**: Ensemble of anomaly detector models
  - Train multiple models independently
  - Decide using an ensemble of anomaly detectors’ **explanations** which records are anomalous





Tasks overview



Data source 1

Anomaly detector SOTA (unsupervised)

Ground truth anomalies

N anomaly detectors (supervised)

XAI model 1

XAI model 1

…

XAI model N

Ensemble to reduce FP

Data source 2

e1, ,e2, e3, e4

Instances

Ground truth anomalies

N anomaly detectors (supervised)

XAI model 1

XAI model 1

…

XAI model N

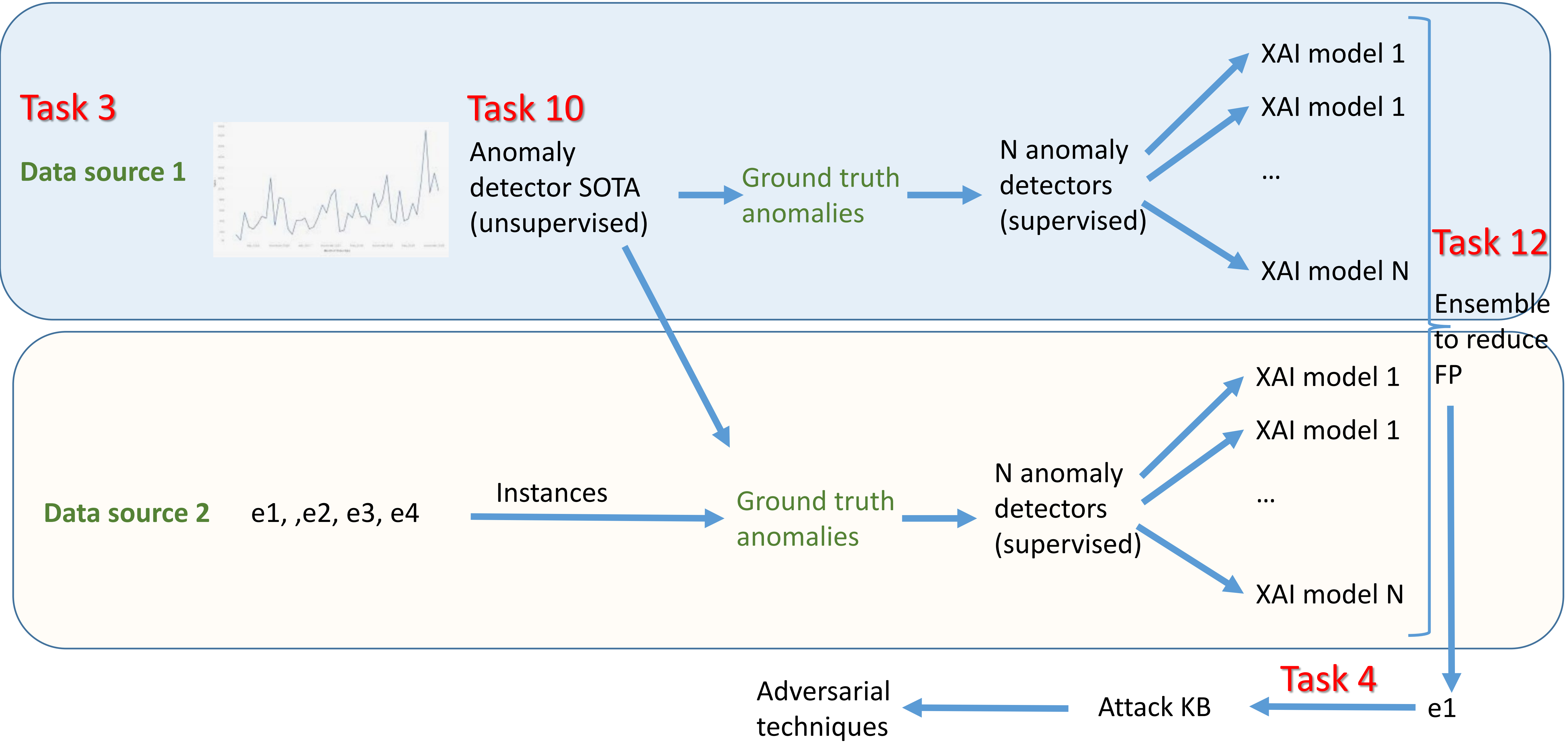
Ensemble to reduce FP

e1

Attack KB

Adversarial techniques

Tasks overview



Task 3

Data source 1

Task 10

Anomaly detector SOTA (unsupervised)

Ground truth anomalies

N anomaly detectors (supervised)

XAI model 1

XAI model 1

…

XAI model N

Task 12

Ensemble to reduce FP

Data source 2

e1, ,e2, e3, e4

Instances

Ground truth anomalies

N anomaly detectors (supervised)

XAI model 1

XAI model 1

…

XAI model N

e1

Task 4

Attack KB

Adversarial techniques

Experiments

- Goals
  - To find a connection in the explanations of anomalies revealed by different models
  - Reduce false positive anomalies using the connection between the explanations
- Challenges
  - The quality of the explanations is limited by the quality of the anomaly detector
  - How to make a decision if an event is anomalous or not – there are many parameters for the decision (number of explaining features, how many models from the ensemble need to agree with each other)

Future work

- Continue exploring different supervised models, as part of the ensemble
- Continue exploring the connection between explaining features
- Experiment on real-world datasets (Arava, etc.)