

Separating between a censorship event and non censorship event

Metadata fields in dataset
stateful_block,server_country...

Measurement Point Approach
Closer to 0 - not a censorship event
closer to 1 - a censorship event

Certain conditions met
within metadata

Certain conditions not
met within metadata

Increase
score

Decrease
score

Final score

A chance that this
dataset is a
censorship event

If score >0.7

A chance that this
dataset is not a
censorship event

If score <0.3

Questions

How reliable is this approach?
We can compare with ground truth or based on
other censorship event reports

How can I utilize decision trees with this approach
so that this strategy is more refined?

Benefits:

Removes false positives

Accurately distinguish between a
event and non event

Reduce amount of false positives
Supports my hypothesis