

5/3/2021

Last day!

Operationalizing data analysis – include in your recommendations

Automate certain tasks

Issues:

Can take a long to build an analysis

Require collaboration

Not being properly scaled

Lack trust/might be incomplete

1. Get the right
2. Accelerate development
3. Ensure collaboration and automation
4. Promote data governance
5. Follow best practices

May require:

1. Experimentation
2. Implementation – alpha/beta tests (monitoring performance after implementation)
3. Expansion
4. Optimize – constant improvement, keep relevant and effective

Model validation:

1. Is the model performant (does it perform better than humans)
Different types of models have different types of tests to assess performance or accuracy
2. Is the model stable? (is it overfit?)
Cross-validation can test this
Some types of models (like decision trees) are inherently unstable
3. Any biases in the model
4. Is the model too sensitive? (with overfitting)
Is the model tolerant to noise?
How does it respond to extreme scenarios?
5. Is the model predictive? – does it actually make good predictions?
6. Is there leakage? – is the accuracy stable over time?