



Pitfalls of Explainable ML :An Industry Perspective

Sahil Verma*, Aditya Lahiri*, John Dickerson, Su-In Lee



Explanations in ML

- Research in explainable ML has taken huge strides
- Yet several problems remain unaddressed, specifically for industry adoption
- By drawing from industrial experience, this work brings light to such problems and poses them as research challenges.

Lack of expressive evaluation metrics

- Variants of explainability technique need to be compared using quantitative metrics
- Current metrics like fidelity and robustness do not capture the usefulness of explanations
- For invigorating future research, we need to standardize a set of accepted metrics

Lack of established workflow for several use cases

- Explainability is required in several stages in a ML model lifecycle
- Most current research focuses only on generating explanations
- The research on taking decisions based on explanations is crucial but is lacking

Lack of scalable implementations

- Most explainability techniques do not scale well to big data
- Their implementation do not support distributed systems like PySpark
- Global explanations obtained by aggregation of local explanations is computationally expensive and time consuming.

Lack of standardized benchmarks

- Most fields have benefited from standardized benchmarks. They allow quantitative comparisons and easy tracking of progress of research.
- Explainability research lacks such standardized benchmarks.
- A part of this can also be attributed to lack of desirable metrics.

Lack of incentives and enforcements

- Ambiguous and non-binding government regulations do not incentivize industry to invest in trustworthy ML infrastructure.
- Even the widely acclaimed GDPR lacks precise language for enforcing right to explanation¹.
- For smaller resource constrained firms, even reputational concerns aren't important enough for investment.

1. Sandra Wachter, Brent Mittelstadt, and Luciano Floridi. 2017. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation.



Thank you!

Q & A

