JENGA - A Framework to Study the Impact of Data Errors on the Predictions of Machine Learning Models

Sebastian Schelter (University of Amsterdam), Tammo Rukat (Amazon), Felix Biessmann (Einstein Center Berlin, Amazon)

`https://github.com/schelterlabs/jenga`

Why Jenga?

Software systems are tested (Unit tests, integration tests, ...)
ML applications are difficult to test

- ML models depend on data
- Real test data is limited, models overparametrized
  - Google's underspecifiation paper (D'Amour et al., 2020)
  - Stochastic Parrots paper (Bender et al., 2021)

What Jenga offers

- Simple API to
- Define Machine Learning tasks
- Corrupt data to test Machine Learning models
- Predefined corruptions for images, text, tables
- Evaluate ML models on tasks
- Implement custom data corruptions

How it works

Apply User Defined Error Generators to labeled data

Missing Values, Scaling, ...

Evaluate Impact on Trained ML Model

labeled test examples

age	income	loan?
18	2,500	no
40	5,000	yes
37	2,800	no

corrupted examples

age	income	loan?
18	**250,000**	no
N/A	5,000	yes
N/A	2,800	no

age	income	loan?
N/A	**-2,500**	no
40	5,000	yes
37	**-2,800**	no