GitHub Project Health Metrics

Tyler Brown brown.tyler@husky.neu.edu

Why does project health matter?

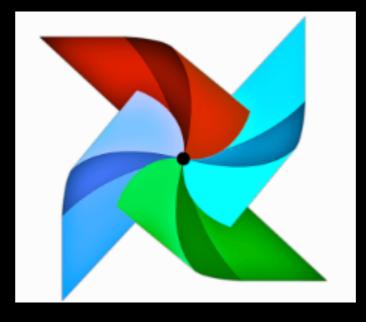
- Maintaining software is time consuming and expensive
- Improving software requires additional resources
- We want to choose Open Source Software that is robust and likely to be maintained

Case Study: Building a Data Pipeline

Asked to build a data pipeline in 2015; two major options



Luigi from Spotify



Airflow from AirBnB

Both projects were being developed on GitHub

Case Study: Building a Data Pipeline

- Luigi and Airflow had very similar features and financial support
- I chose Luigi to build our data pipeline in 2015
- Airflow joined the Apache Foundation's Incubator program in March 2016

Case Study: Building a Data Pipeline

March 13, 2019	Luigi	Airflow
Commits	3,755	6,074
Releases	45	105
Contributors	392	742
Stars	11,096	11,425

Case Study: Building a Data Pipeline

- Choosing Airflow would have been better
 - Technical debt to switch pipelines
 - Responsiveness to new features
 - Fortunately, both projects are currently maintained

How can we measure project health?

- Establish a baseline using projects we consider mature and healthy like Apache Lucene
 - Frequency of releases
 - Commits per release
 - Contributors per release
- Compare other projects on GitHub to this baseline

How can we predict project health?

- Requirements
 - Models need to be interpretable
 - Causality instead of Correlation
 - Probabilities instead of classification
- Objective
 - Probability of meeting or exceeding the baseline

Implications for OSS

- Choose healthy OSS to build healthy software
 - Cyber Security
 - Technical Debt
 - Access to new features

Why should you care?

- We saw how I chose Luigi instead of Airflow for a data pipeline framework in 2015. Airflow became the clear winner in 2019.
- We saw how actively maintained and mature projects can be used as a baseline for project health
- We saw how unhealthy projects struggle with cyber security, lead to technical debt in your stack, and struggle to provide you with access to new features

Questions?

• Email: <u>brown.tyler@husky.neu.edu</u>

• GitHub: tbonza