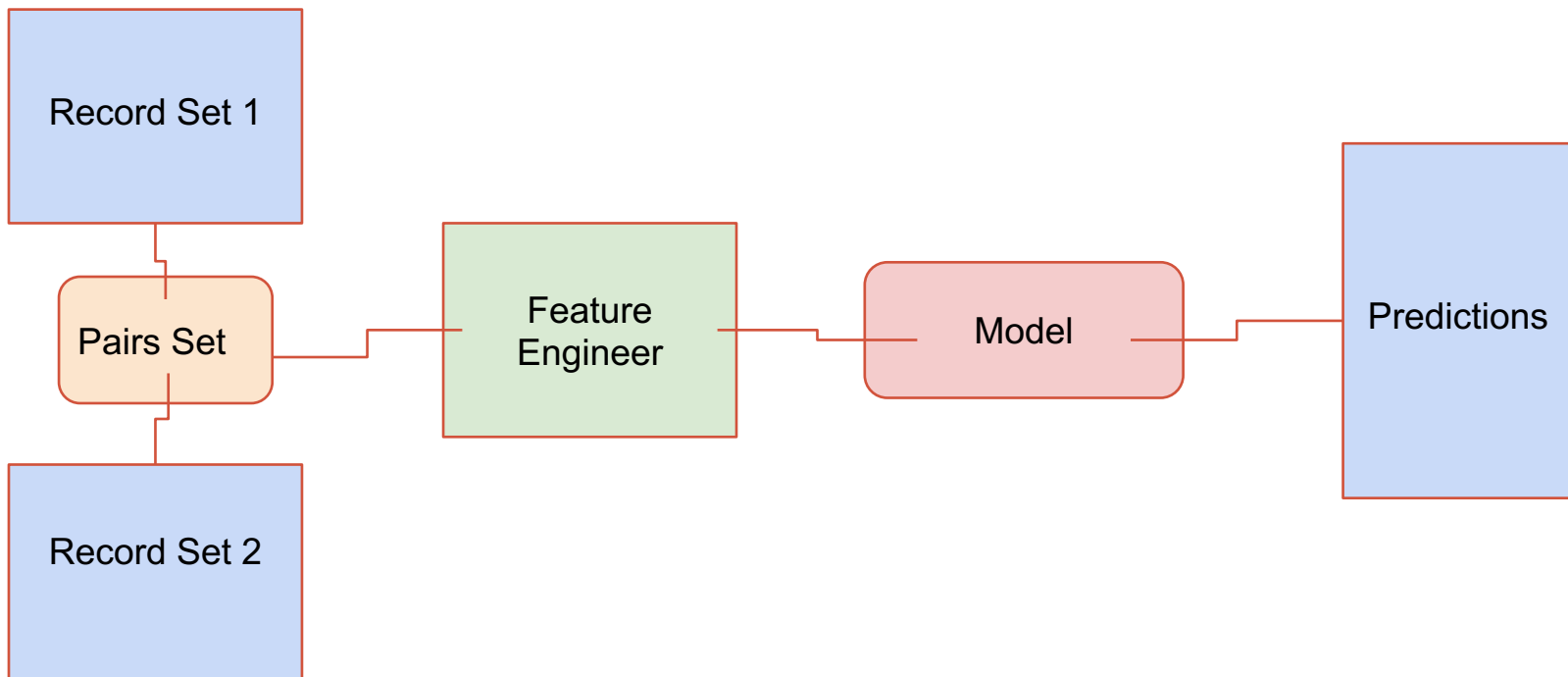# Splycer

Ben Busath

bbusath5@byu.edu

What is Splycer?

- Splycer is a Python package built by BYU students for record linkage
- Includes all tools necessary to train and run machine learning models for linking people across record sets
- Publically available on GitHub

What does Splycer do?

Chester A Molland
United States Census, 1920

	Tools

Name:	Chester A Molland • Edit
Event Type:	Census
Event Date:	1920
Event Place:	Sterling, Vernon, Wisconsin, United States
Gender:	Male
Age:	1
Marital Status:	Single
Race:	White
Race (Original):	White
Relationship to Head of Household:	Son
Relationship to Head of Household (Original):	Son
Birth Year (Estimated):	1919
Birthplace:	Wisconsin
Father's Birthplace:	Wisconsin
Mother's Birthplace:	Wisconsin
Sheet Letter:	A

Chester A Molland
United States Census, 1930

	Tools

Name:	Chester A Molland
Event Type:	Census
Event Date:	1930
Event Place:	Sterling, Vernon, Wisconsin, United States
Gender:	Male
Age:	11
Marital Status:	Single
Race:	White
Race (Original):	White
Relationship to Head of Household:	Son
Relationship to Head of Household (Original):	Son
Birth Year (Estimated):	1919
Birthplace:	Wisconsin
Father's Birthplace:	Wisconsin
Mother's Birthplace:	Wisconsin

.998 match score

Splycer Features

- Works with any SciKit-Learn classifier
- Robust feature engineering
- User can choose internal data structures
- Fast Blocking
- Compiled in Cython
- Can Save/Load models, feature engineers

Splycer Applications

- Through Splycer, we were able to link the entire 1930 - 1940 censuses in just under twelve hours!
- Other applications:
  - Entity Resolution
  - Research
  - Medical Data
  - Database Administration