
Data processing pipelines for the people Documentation

Release 0.1

Gabriele Lanaro

October 31, 2016

Building a machine learning pipeline involves exploring different models, pre and post-processing, feature selection and data cleaning, **ezpipe** makes the process of wiring up all those different steps in an easy, efficient and reproducible way.

Here is an example on how to setup a pipeline for sentiment analysis.

```
from ezpipe import Pipeline
p = Pipeline()
p.add_map('X', 'tokenize', tokenize)
p.add_transformer('tokenize', 'vectorized', CountVectorizer(), name='vec')
p.add_model('vectorized', 'sentiment', LogisticRegression(), name='reg')
```

Train the pipeline.

```
X = ['Good boy',
     'Bad boy']
y = [1, 0]
p.fit('vec', X=X)
p.fit('reg', X=X, sentiment=y)
```

Make predictions

```
p.get('sentiment', X=['Good girl', 'Bad girl'])
# [1, 0]
```

Contents: