

Using Wikipedia Data as Political Polls

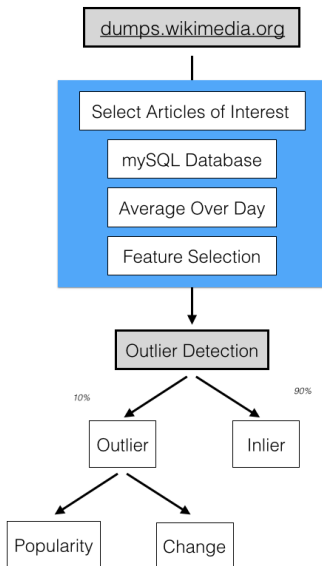
Abram Van Der Geest

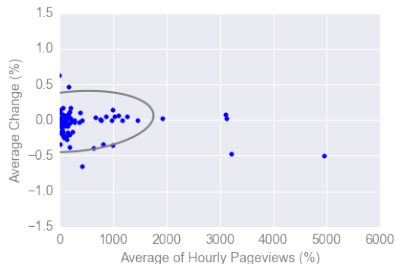
Insight Data Science

abram.ghost@gmail.com

September 25, 2015

Pipeline





- Extract time slice
- Select features
 - ▶ Average views
 - ▶ Average rate of change
- Outlier Detection
- EllipticEnvelope (sklearn)
 - ▶ Gaussian distribution
 - ▶ 10% contamination

Data Story / Validation

