EDCT GE2550: DATA SCIENCE IN EDUCATION

Big Data, Learning Analytics & The Information Age

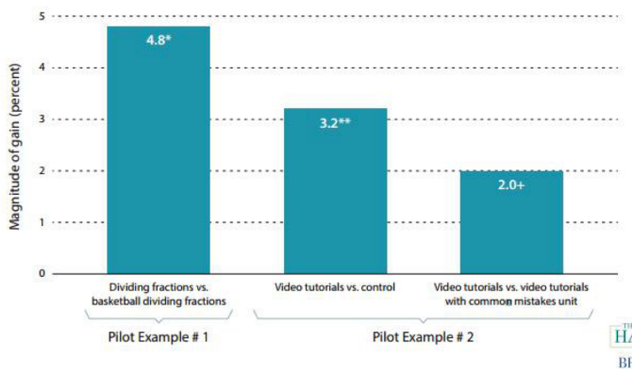
4/8/16 10:23 AM

In the news

Understanding the Origins of Ed-Tech Snake Oil

By Michael Feldstein | APRIL 06, 2016

FIGURE 1.
EDUSTAR Results for Two Pilot Examples



THE CHRONICLE of Higher Education

How Ever-Worsening Malware Attacks Threaten Student Data

Emerging cybercrime trends, such as malvertising and ransomware, are creating significant challenges for school and district IT staff.

EduStar Platform Promises Quick, Randomized Ed-Tech Trials

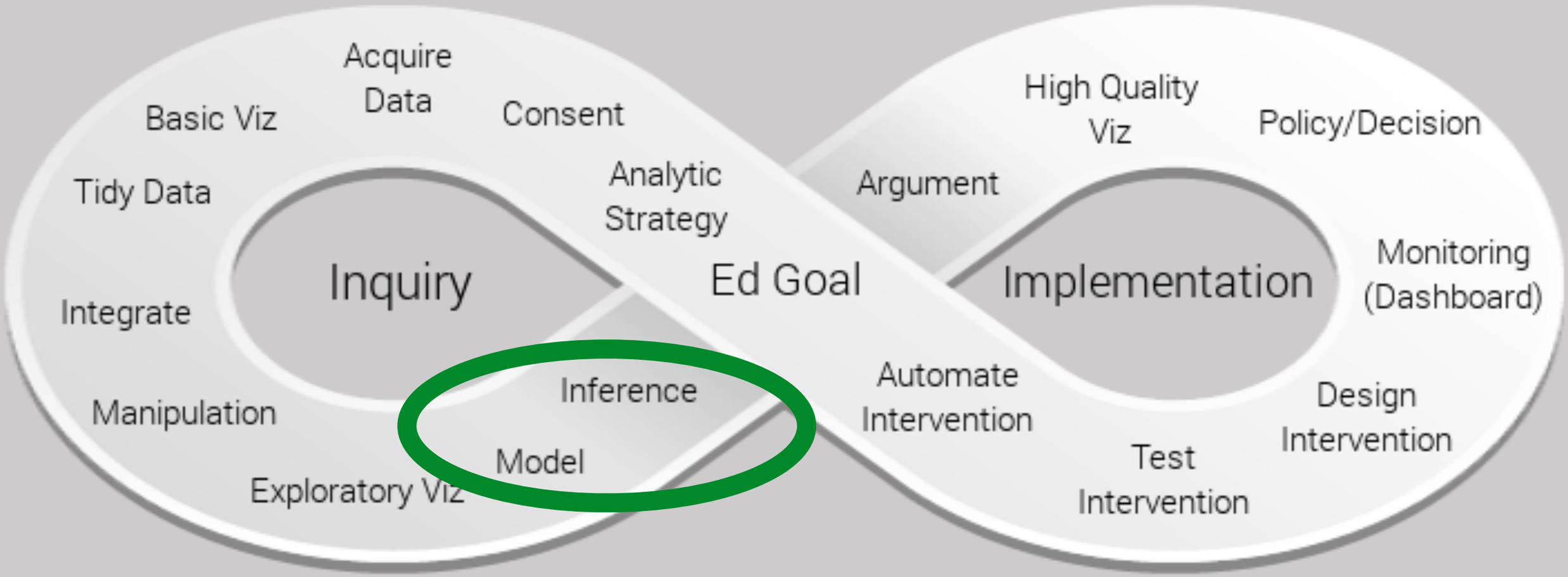
Rise of the Fembots: Why Artificial Intelligence Is Often Female

by Tanya Lewis, Staff Writer | February 19, 2015 07:15am ET

Today

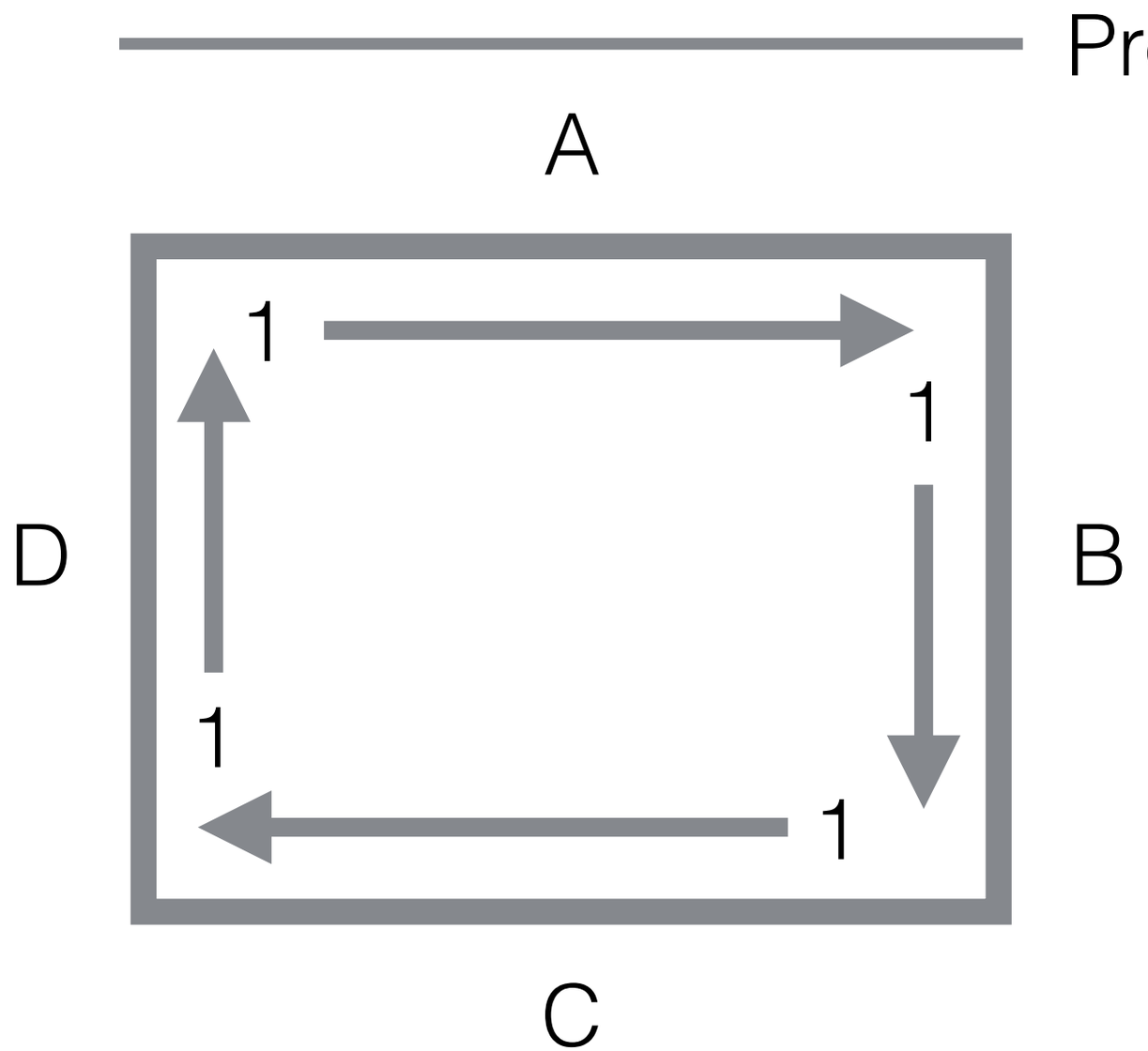
In the news	6:45 - 6:55
Quiz	6:55 - 7:05
Natural Language Processing	7:05 - 7:15
NLP Activity	7:15 - 7:30
Twitter Activity	7:30 - 7:45
Random Forests	7:45 - 8:00
Twitter	8:00 - 8:20

Ed Data Science Cycle



Quiz

http://bit.ly/1WdDZeR



Final Assignment

- Next week I will ask you to choose:
  - Data you will use
  - Method you will apply

Opportunities

- Summer internship

- NSF Big Data Hub Internships

Natural Language Processing

4/8/16 10:23 AM

NLP

Analyses of language produced by humans (by computers)

- Treats language as a varied pool of information sources
- In order to:
  - Understand language (Cognitive Science)
  - Respond to the speaker appropriately (AI)

- Examples
  - Translation
  - Automated feedback (education, shopping)
  - Study linguistics, cognition, development, etc.

Methodological History

1930s

Understanding

Rule based

- Complex sets of rules (grammar/syntax)
- Chomsky

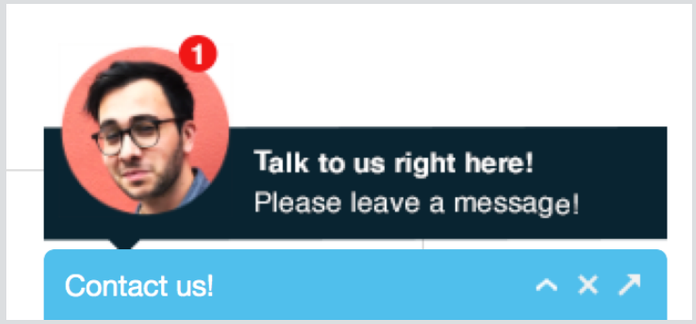
1980s

Processing

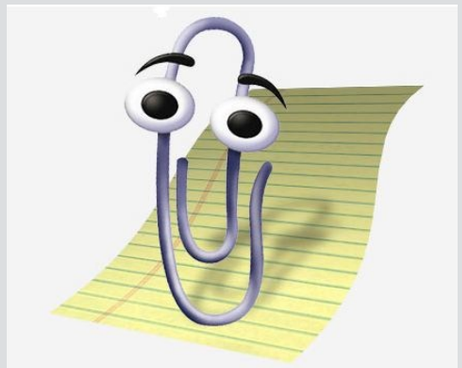
Statistical

- Infer rules from data
- IBM

Industry



Education

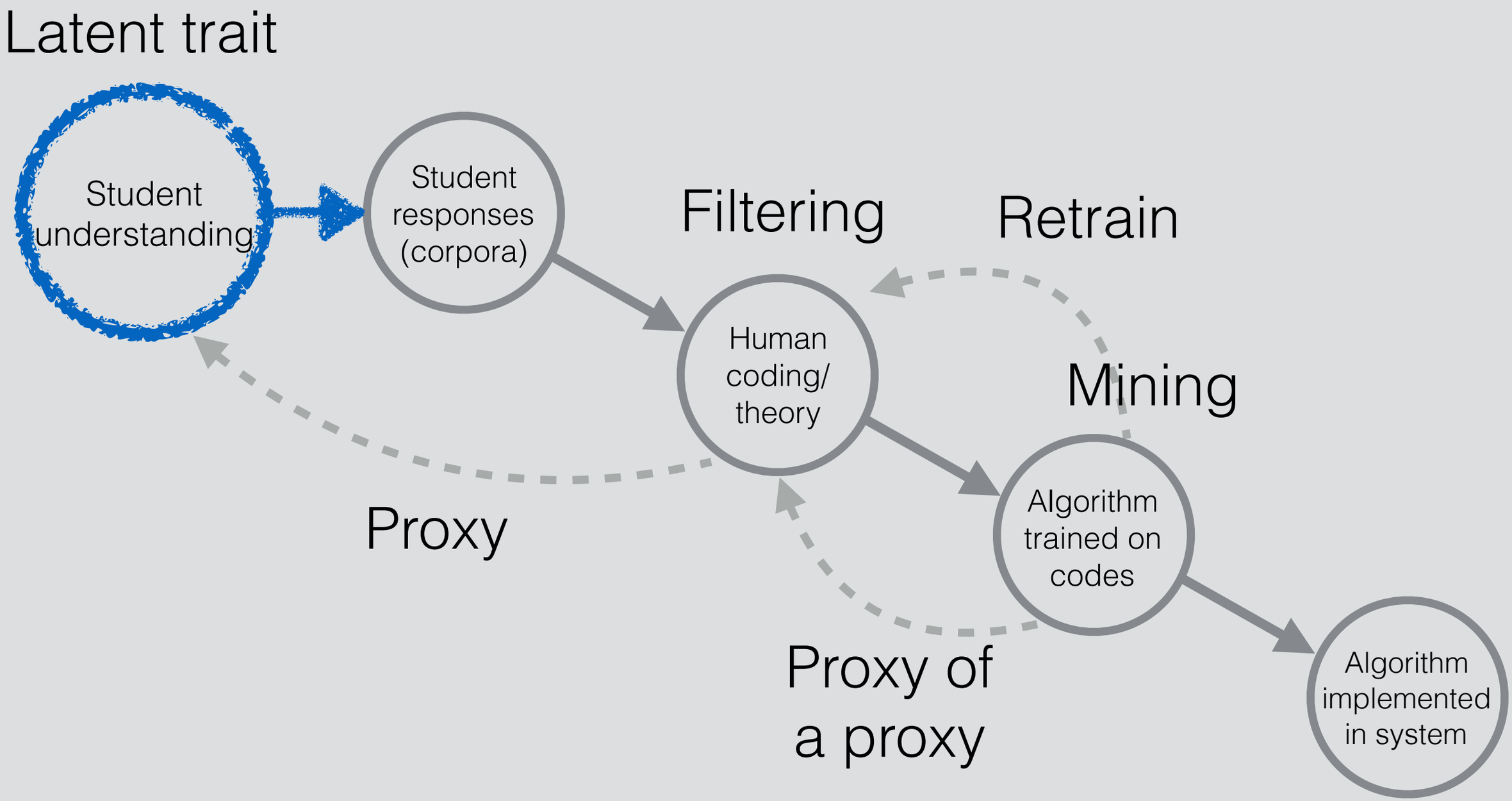


Essential Problem

- Heterogeneity
- We get rid of this by asking MCQ questions - but we also throw out a lot of information when we do that

- Collect more data and more complex data through written answers

Overall Method



Coding

Word counting

Google books Ngram Viewer

Types of Expressions

"I don't know…"
"I dunno…"

Stemming

Take the root of the word:
educate, education, educating

Tokenization (bag of words)

Chopping word/phrase into *tokens*

- Remove punctuation
- Find best number of letters to represent a word/meaning
- Consider all possible versions of word
- Stop word removal

↓

Features

Algorithms

Feature selection

- Not all tokens are useful, which ones can we scrap?

Feature extraction

- Extracting features from combining tokens