Training Evaluation Models

FEATURE ENGINEERING

David Epstein
Senior Data Scientist, Socure

Open Data Science Conference, Boston, MA
May 30-31, 2015
#ODSC, @opendatasci

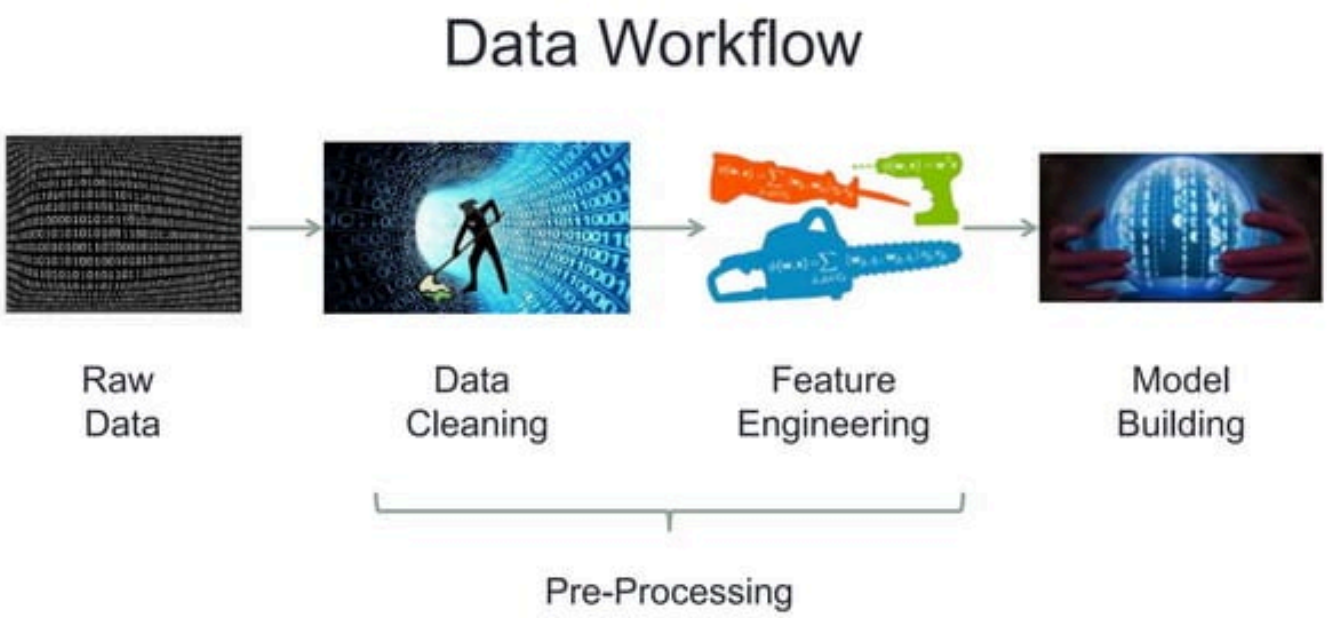
Talk Outline

- What is feature engineering?
- Limits on number of features
- How to select a “good” set of features
- Standard FE techniques

- TL;DR: As we get better and better models, focus shifts to what we put into them
- FE interacts with other key areas of DS

Feature Engineering

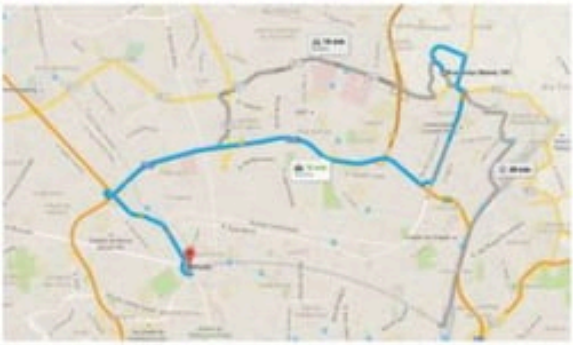
- (My) Definition: Transforming data to create model inputs.



Examples from Kaggle Competitions

Netflix

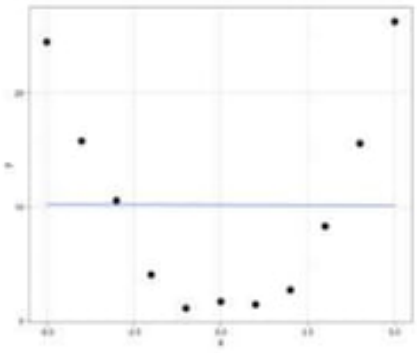
Titanic



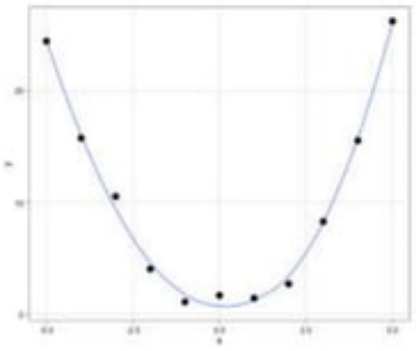
Portuguese Taxis

"Golden Features"

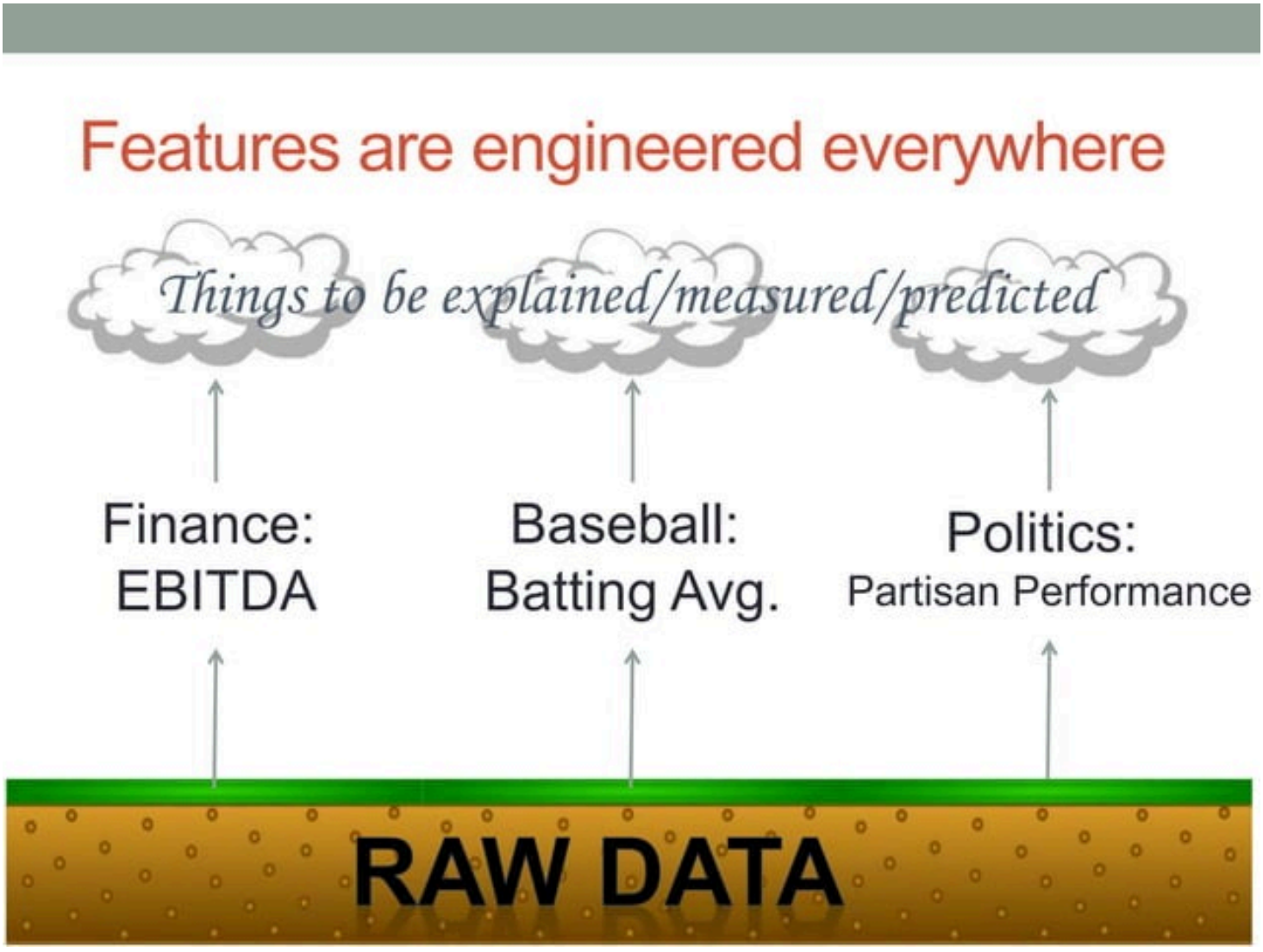
How does it work?



```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.17912    2.92472    3.48  0.00693 **
x           -0.00923    0.92488   -0.01  0.99225
---
Residual standard error: 9.7 on 9 degrees of freedom
Multiple R-squared: 1.107e-05, Adjusted R-squared: -0.11
F-statistic: 9.96e-05 on 1 and 9 DF,  p-value: 0.9923
```



```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.29879    0.48994    0.61    0.559
x           -0.00923    0.10256   -0.09    0.930
x2           0.98803    0.03672   26.91 3.92e-09 ***
---
Residual standard error: 1.076 on 8 degrees of freedom
Multiple R-squared: 0.9891, Adjusted R-squared: 0.9863
F-statistic: 362 on 2 and 8 DF, p-value: 1.427e-08
```


Features are engineered everywhere
Things to be explained/measured/predicted
Finance:
EBITDA
Baseball:
Batting Avg.
Politics:
Partisan Performance
RAW DATA

The Big Questions

- Seen in this light, FE is ubiquitous (as all truly important concepts are)
- Any time you construct an intermediate variable, you're doing FE
- Two questions naturally arise:
  1. How do you construct "good" features?
  2. What are the limits on this process?
- I'll answer the second one first, because it's easier….

Limits on Feature Engineering

- In medical studies, social science, financial analysis, etc., two main problems emerge
- Eating up degrees of freedom: relatively small data sets
  - # of respondents in survey
  - # of patients in trial
  - # of elections to Congress
  - If your data lives in an *NxK* matrix, you want to make sure that *K* is small relative to *N*
- Relevance to hypothesis testing, emphasis on explanation
  - You generally start with an equation defining the relationship between the key independent and dependent variables
  - Other variables enter your model as controls, not really interested in their functional form

Limits on Feature Engineering

- In most modern data science applications, neither is an issue
  - We start with lots of data, and
  - Care more about prediction than explanation
- So why not add in lots of extra variables?
  - Think of your data not as what goes into your model, but a starting point for the cr n then be combined…

Various Evaluation Models

1. Kirkpatrick's Model
2. Jack J Phillips Training Model
3. CIPP Model
4. COMA Model

Kirkpatrick Model

- The Four Levels of Evaluation, also referred to as the Kirkpatrick Evaluation Model, was created by Donald Kirkpatrick, Ph.D. to define the four levels of training evaluation. The four levels of evaluation are:
- **Level 1: Reaction** measures how participants react to the training (e.g., satisfaction).
- **Level 2: Learning** analyzes if they truly understood the training (e.g., increase in knowledge, skills or experience).
- **Level 3 :Behavior** looks at if they are utilizing what they learned at work (e.g., change in behaviors).
- **Level 4 :Results** determines if the material had a positive impact on the business / organization.

The Four Levels

THE KIRKPATRICK MODEL

Phillips ROI Model

- One of the most frequently quoted aspects of Phillips' model is the addition of a fifth level of evaluation to Kirkpatrick's Learning Evaluation Model which is return on investment (ROI).
- Phillip's model states that after determining a learning program's business impact at Kirkpatrick's Level 4, we can translate that impact into monetary terms and compare it to the total cost of the program to calculate ROI. These costs include program development and delivery, plus the labor cost of time for learners to complete the training.


ROI
Did the training investment provide a positive return on investment?
Results
Did the training have a measurable impact on performance?
Impact
Did the learners behavior change as a result of the training?
Learning
Did knowledge transfer occur?
Satisfaction
Did the learners enjoy the training?

Calculation of ROI

- The training ROI formula as follows:

  ROI(%)=NET Program Benefits/Program cost x100

CIPP Model

- CIPP evaluation model is a Program evaluation model which was developed by Daniel Stufflebeam and colleagues in the 1960s.
- CIPP is an acronym for Context, Input, Process and Product.
- CIPP is an evaluation model that requires the evaluation of context, input, process and product in judging a program me's value.
- CIPP is a decision-focused approach to evaluation and emphasizes the systematic provision of information for program management and operation.

Four aspects of CIPP Model

The four aspects are:

- Context
- inputs
- process
- product
- These four aspects of CIPP evaluation assist a decision-maker to answer four basic questions:

- **Context**: What should we do?

  This involves collecting and analyzing needs assessment data to determine goals, priorities and objectives.

- **Inputs**: How should we do it?

  This involves the steps and resources needed to meet the new goals and objectives and might include identifying successful external programs and materials as well as gathering information.

- **Process** :Are we doing it as planned?

  This provides decision-makers with information about how well the program me is being implemented. By continuously monitoring the program, decision-makers learn such things as how well it is following the plans and guidelines, conflicts arising, staff support and morale, strengths and weaknesses of materials, delivery and budgeting problems.[

- **Product**: Did the program me work?

  By measuring the actual outcomes and comparing them to the anticipated outcomes, decision-makers are better able to decide if the program should be continued, modified, or dropped altogether. This is the essence of product evaluation.