
H2O
WORLD
2017

Repurposing data to solve emerging problems

Arturo Castellanos, PhD

Assistant Professor - Baruch College (CUNY)

www.arturocastellanos.com

H2O WORLD 2017

In the course of normal business, organizations generate electronic records describing daily operations and transactions.

IDC, estimates that more than **90%** of the enterprise data generated are **unstructured**

Based on existing data, could we raise red flags to avoid these events from happening?

Suicide of foster child, 7, prompts review of DCF drug policy

By Kris Hundley, Times Staff Writer
Friday, May 8, 2009 9:13pm

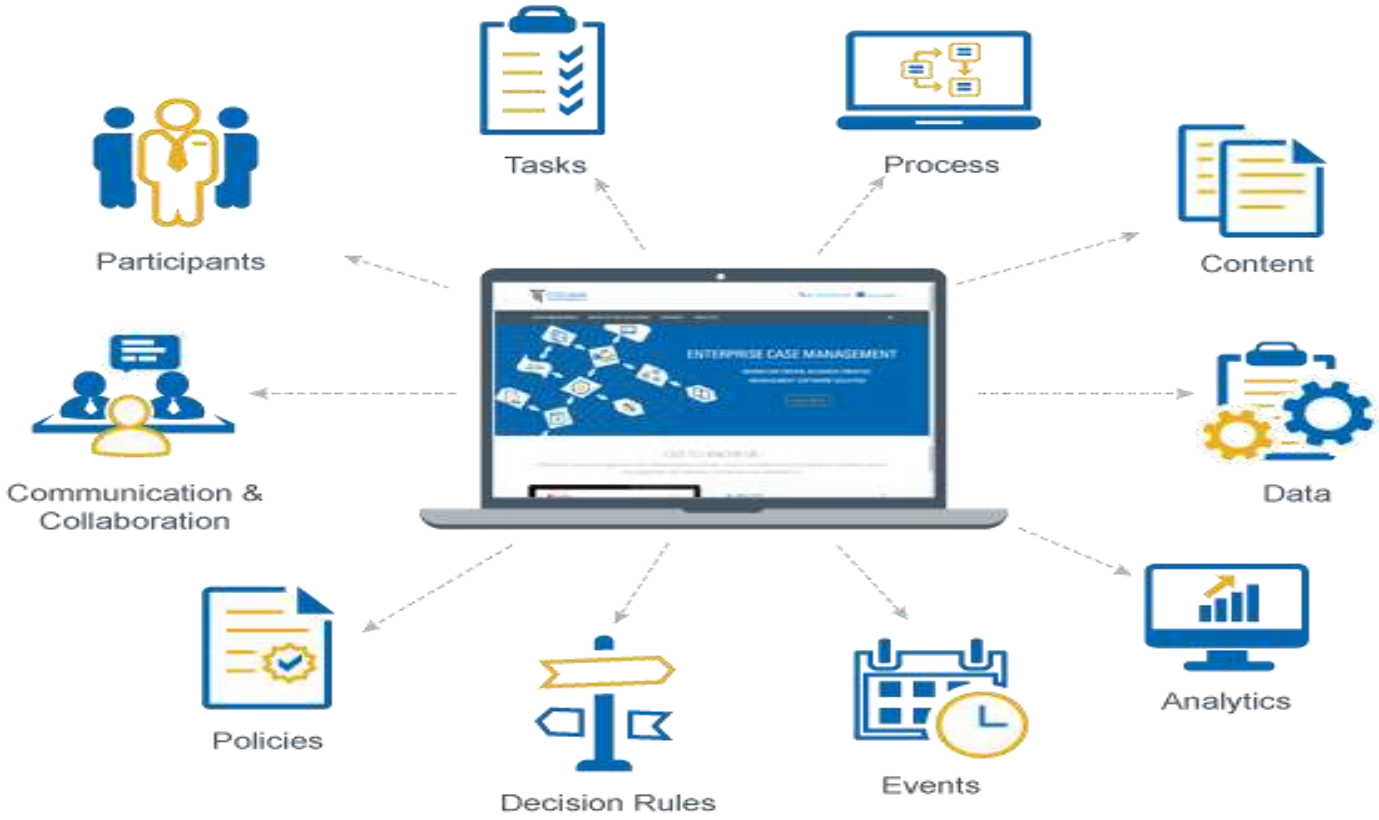
22 Tweet 22 41

Last month, Gabriel Myers, age 7, died by suicide in his South Florida foster home, hanging himself on an extendable shower hose.

Image source: https://pixabay.com/en/worried-girl-woman-waiting-sitting-413690/

Case Management is **complex**

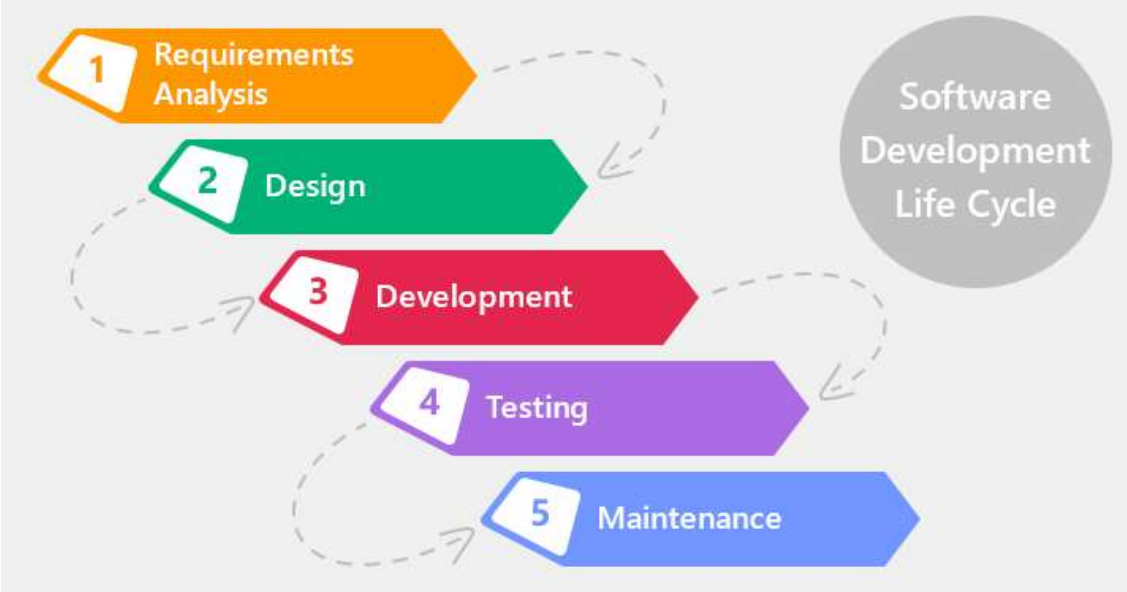
Process-driven. Human-in-the-loop



One organization and three case management agencies

Image source: https://www.columncase.com/what-is-case-management.html

Information Systems design: often static.. Yet, business problems evolve



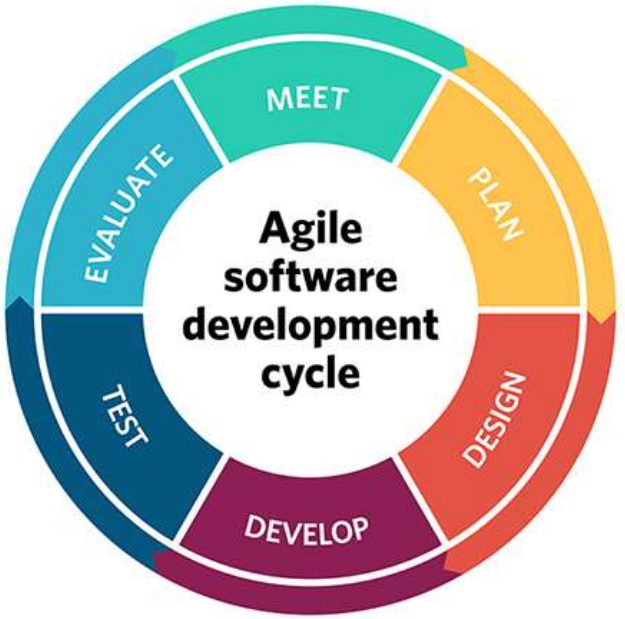
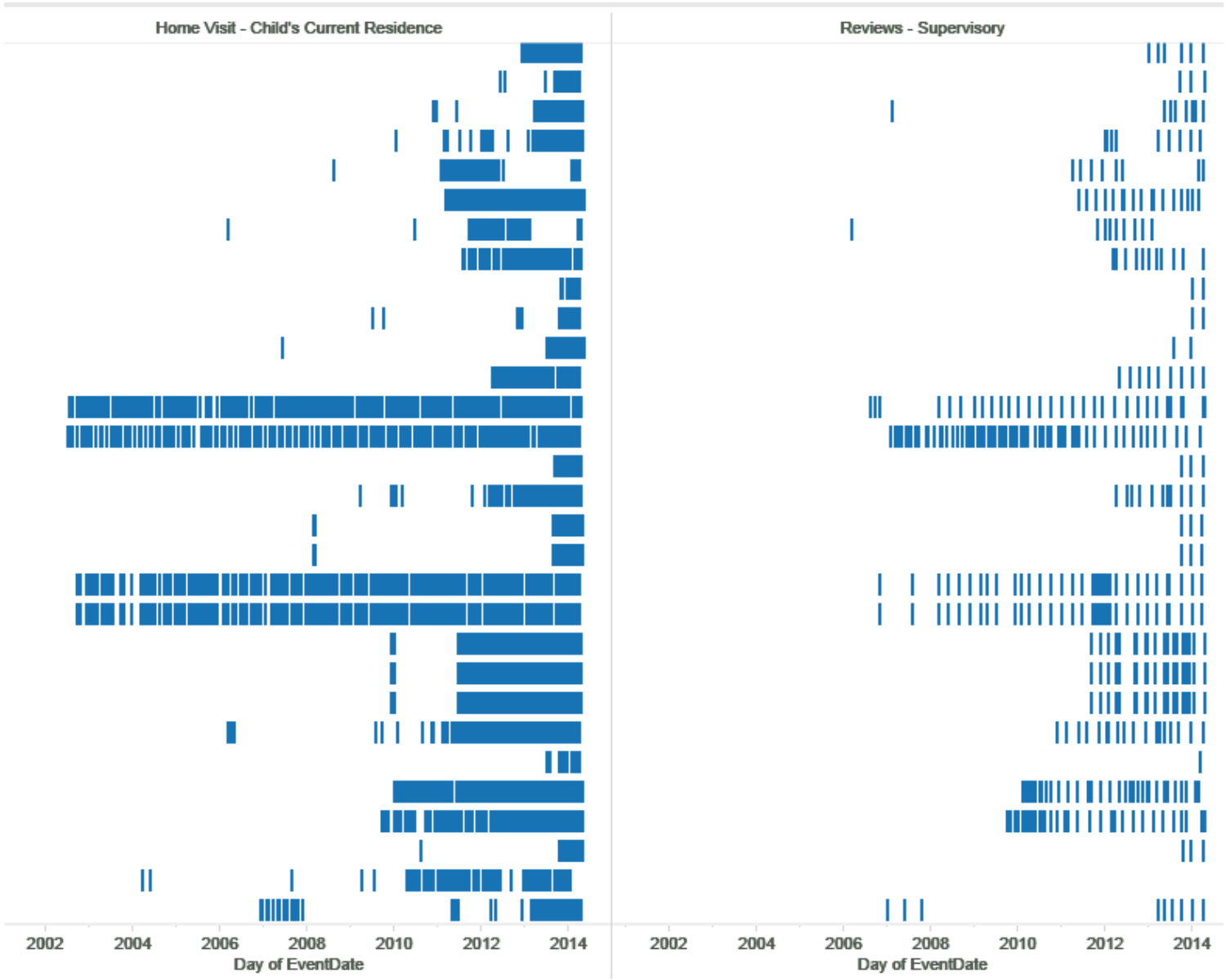


Image Source: https://xbsoftware.com/blog/software-development-life-cycle-waterfall-model/
http://searchsoftwarequality.techtarget.com/definition/agile-software-development



Children in foster care are **three to ten times** more likely to suffer from **mental health conditions**[1].

Psychotropic medication is often prescribed to **help them cope with behavioral problems**

No business process in place to specifically track medication intake by foster children!

1. (Harman, Childs, & Kelleher, 2000)

H2O WORLD 2017

Can we identify kids on psychotropic drugs based on their home visit case notes?

1. (Harman, Childs, & Kelleher, 2000)

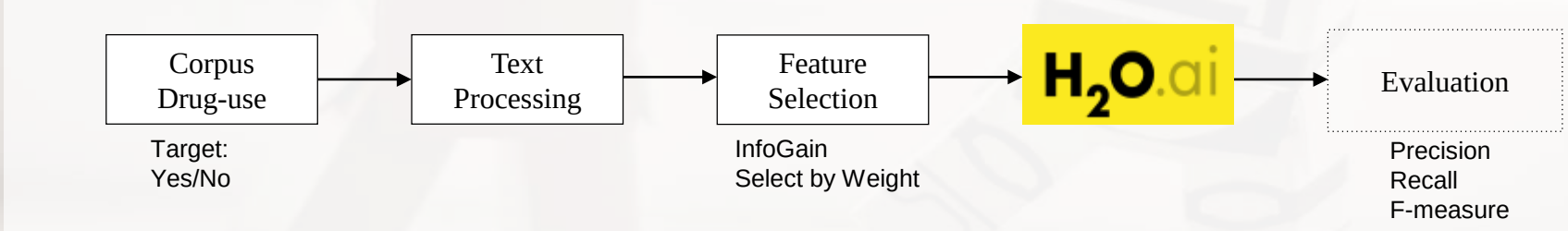
H2O WORLD 2017

Easy, We know how to solve this, right?

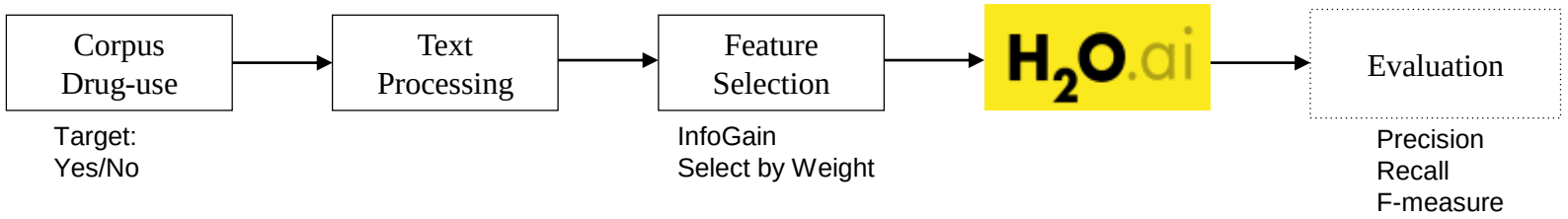
Supervised learning (gold standard): case note + label

Features: Text

Target: binary - med intake



H2O WORLD 2017



Works well, let's apply the model to the data from the other two agencies

Predicting cases of psychotropic drug use

FCMA	Precision (%)	Recall (%)	F-measure (%)
FCMA A	71.93	80.39	75.93
FCMA B	85.71	87.8	86.75
FCMA C	76.92	81.08	78.95

All things equal, **why the difference in performance?**

H2O WORLD 2017

Institutional Theory:

Organizational activity can become **a pattern that is repeated** by individuals in the organization.

Institutional theory considers the processes by which structures, including schemes, **rules, norms, and routines** **become established** as authoritative guidelines for social behavior[1]

What makes an agency perform better?

1. W.R. Scott, 1987

Psychology theory

Cognitive research plays an important role in the development of information systems since **these systems are <u>representations</u>** of the real-world[1]

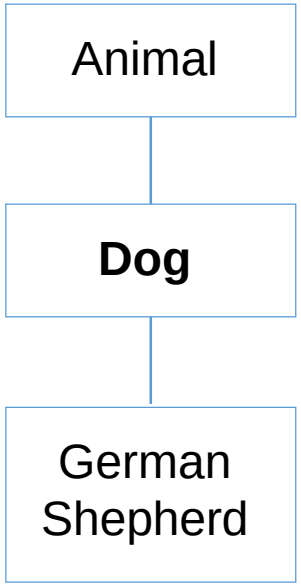
Concepts provide two primary functions[2]:

a. **Cognitive economy:** maximally abstracting from individual differences among objects and then grouping objects in categories of larger scope

b. **Inferential utility:** The ability to predict attributes of instances of a class, increases as the scope of the class decreases (more specific)

1. Davern, Shaft, & Te'eni, 2012
2. Lakoff, 1987; Eleanor Rosch, 1999; Smith & Medin, 1981

Consider the following hierarchy. Which one provides:



a. **Cognitive economy:** maximally abstracting from individual differences among objects and then grouping objects in categories of larger scope

b. **Inferential utility:** The ability to predict attributes of instances of a class, increases as the scope of the class decreases.

a. Higher cognitive economy?
b. Which one provides higher inferential utility?

Why is this important?

- The more specific the more inferences we can make.

Based on an excerpt from the note, which of the following children is taking psychotropic medication?

Child 1. "…Takes 5 mg of vyvanse by mouth once a day… she has to call the doctor to schedule for a refill…"

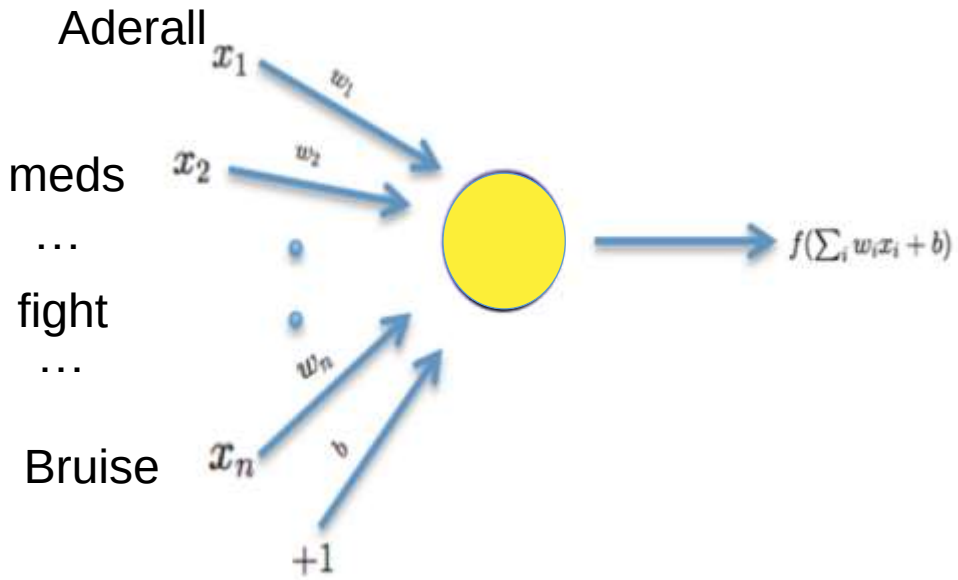
Child 2. **"…**child has an immune system medical condition that requires many medications to keep her healthy…"

For unstructured data, higher the specificity is preferable

For structured data… It depends! There are tradeoffs!

Why H2O?

- Multi-layer feed-forward architecture
- Fast and memory-efficient (scalability)
- Performs incredibly well**, out-of-the-box!**
- User-friendliness: from GUI to open source code
- Great at not overfitting (big issue with other tools!)
- Ability to run on a one-node local cluster, which forces reproducibility



H2O WORLD 2017

Thank You

arturo.castellanos@baruch.cuny.edu
www.arturocastellanos.com