Building a model is the *least* important part of your job

Kimberly Shenk
Director, Data Science Solutions
Domino Data Lab

DOMINO

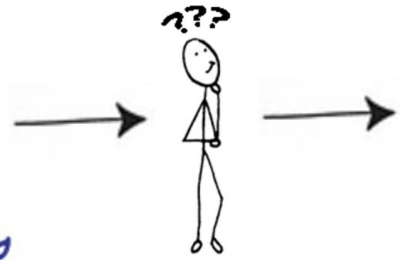
We are told a beautiful story


DATA
DATA DATA
DATA DATA
DATA DATA
DATA DATA DATA DATA
DATA DATA
DATA DATA DATA
DATA DATA DATA
DATA DATA DATA DATA
DATA DATA DATA DATA
DATA DATA DATA DATA DATA DATA
DATA DATA DATA DATA

DOMINO


DATA
DATA DATA DATA
DATA
DATA DATA
DATA DATA DATA DATA
DATA DATA
DATA DATA DATA
DATA DATA DATA DATA
DATA DATA DATA DATA
DATA DATA DATA DATA DATA DATA
DATA DATA DATA DATA
???

DOMINO



Algorithm 1 Boosted K Nearest Neighbour

Inputs:
  $S = s_i = (x_i, y_i)$
Initialize:
  $w_i^0 \leftarrow 0,\ i = 1, \ldots, n$
  $S_0 \leftarrow S$
for t = 1 to T **do**
  $S_t \leftarrow S_{t-1}$
  for $s_q \in S_t$ **do**
    $N_q \leftarrow$ k nearest neighbors
    of s_q using $D(s_q, s_i)$
    label(s_q)= $argmax \sum_{s_i \in N_q} D(s_q, s_i)$;
    if label(s_q)$\neq y_q$ **then**
      for $s_i \in N_i$ **do**
        if $y_i \neq y_q$ **then**
          $w_i^t \leftarrow w_i^t - \lambda / d(x_q, x_i)$;
        else
          $w_i^t \leftarrow w_i^t + \lambda / d(x_q, x_i)$;
        end if
      end for
    end if
  end for
  if label(s_q)= $y_q \forall_{s_q}$ **then**
    break
  end if
end for

DOMINO

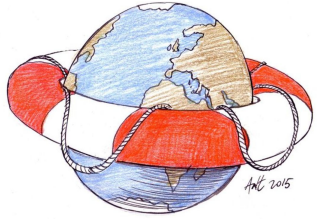
Algorithm 1 Boosted K Nearest Neighbour

Inputs:
  $S = s_i = (x_i, y_i)$
Initialize:
  $w_i^0 \leftarrow 0,\ i = 1, \ldots, n$
  $S_0 \leftarrow S$
for t = 1 to T **do**
  $S_t \leftarrow S_{t-1}$
  for $s_q \in S_t$ **do**
    $N_q \leftarrow$ k nearest neighbors
    of s_q using $D(s_q, s_i)$
    label(s_q)= $argmax \sum_{s_i \in N_q} D(s_q, s_i)$;
    if label(s_q)$\neq y_q$ **then**
      for $s_i \in N_i$ **do**
        if $y_i \neq y_q$ **then**
          $w_i^t \leftarrow w_i^t - \lambda / d(x_q, x_i)$;
        else
          $w_i^t \leftarrow w_i^t + \lambda / d(x_q, x_i)$;
        end if
      end for
    end if
  end for
  if label(s_q)= $y_q \forall_{s_q}$ **then**
    break
  end if
end for

DOMINO

Algorithm 1 Boosted K Nearest Neighbour

Inputs:
  $S = s_i = (x_i, y_i)$
Initialize:
  $w_i^0 \leftarrow 0,\ i = 1, \ldots, n$
  $S_0 \leftarrow S$
for t = 1 to T **do**
  $S_t \leftarrow S_{t-1}$
  for $s_q \in S_t$ **do**
    $N_q \leftarrow$ k nearest neighbors
    of s_q using $D(s_q, s_i)$
    label(s_q)= $argmax \sum_{s_i \in N_q} D(s_q, s_i)$;
    if label(s_q)$\neq y_q$ **then**
      for $s_i \in N_i$ **do**
        if $y_i \neq y_q$ **then**
          $w_i^t \leftarrow w_i^t - \lambda / d(x_q, x_i)$;
        else
          $w_i^t \leftarrow w_i^t + \lambda / d(x_q, x_i)$;
        end if
      end for
    end if
  end for
  if label(s_q)= $y_q \forall_{s_q}$ **then**
    break
  end if
end for



DOMINO

But then we grow up

A beautiful Model on the Shelf

A beautiful Model on the Shelf

=

DOMINO

A beautiful Model on the Shelf

=

DOMINO

Why does this happen?

Why does this happen?

Businesses know they need Data Science

DOMINO

Why does this happen?

Businesses know they need Data Science

but that doesn't mean they know how to use it

DOMINO

Why does this happen?

Businesses know they need Data Science

but that doesn't mean they know how to use it

That's why they hired you...

Stop modeling

Stop modeling… so much

Stop modeling… so much

1. You are in the **Sales** business
2. Learn to be an **Interpreter**
3. Think like a **CEO**
4. **Data Engineers** are your best friend

You are now in the Sales business

You have to convince people in the organization that you can bring value

- Most people don’t know how to leverage your skills
- Craft compelling pitches with real examples
- Believe in what you are selling

You are an Interpreter

You don’t want to deliver things that aren’t actually useful

- Exercise the 5 “Whys”
- Many times what is asked for won’t really solve the problem
- You have to decipher what the problem *really* is

Think like a CEO

Make sure what you are working on will actually drive the business forward

- Learn how to say “No” - practice and role play!
- What decision will be made based on the work you are doing?
- What might change based on your findings?
- Be able to assess what is a high priority for the business as A WHOLE

Data Engineers are your best friends

You actually can't do your job without them

- Walk them through your day-to-day, literally
- Tie your work back to business objectives - your work isn't just theoretical
- Bring in your business partners, especially when integrating with tools or operations

Tips?

.07%*

**probability of being a unicorn*

https://techcrunch.com/2013/11/02/welcome-to-the-unicorn-club/

Stop looking for Unicorns

They will cost you too much

It’s HARD to keep them happy

They take forever to find (do they *really* even exist?)

DOMINO

It’s hard to “Do All” and “Be All” Things

The whole is greater than the sum of its parts” - Aristotle

- Which area do people on your team naturally gravitate to?

sales pitchers - problem interpreters - mini-strategists/ CEO thinkers - data Eng socialites

- Build out individual expertise - give each person one area to focus on
- Divide and Conquer

Thank you

Kimberly Shenk
kimberly@dominodatalab.com

DOMINO