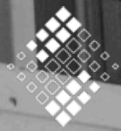
Building the World's Largest Database of Car Features from PDFs



SILICON VALLEY DATA SCIENCE

John Akred
Chief Technology Officer
Silicon Valley Data Science

Robert Munro
Chief Executive Officer
Idibon

IDIBON

JOHN AKRED
@BigDataAnalysis

Founder & CTO

Silicon Valley Data Science

Consulting firm of elite data science and engineering teams who specialize in data-driven product development and business transformation.

SILICON VALLEY DATA SCIENCE

Robert Munro
@WWRob

Founder & CEO

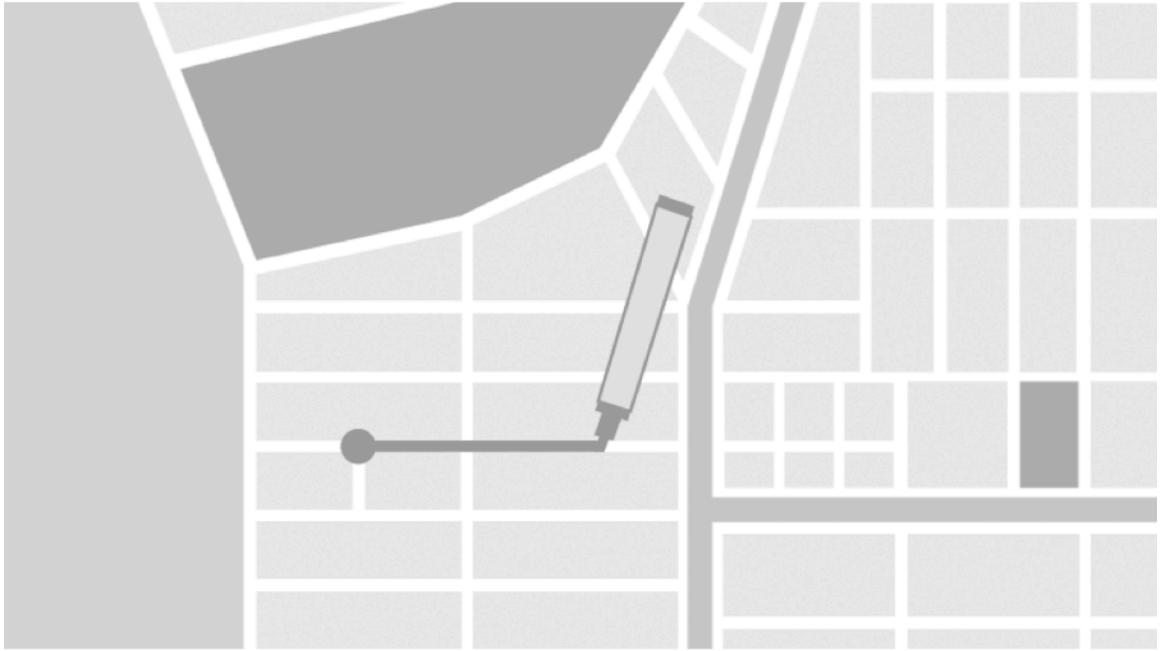
Idibon

Developers of cloud-based natural language processing services that adapt to business-specific problems, in any language.

IDIBON

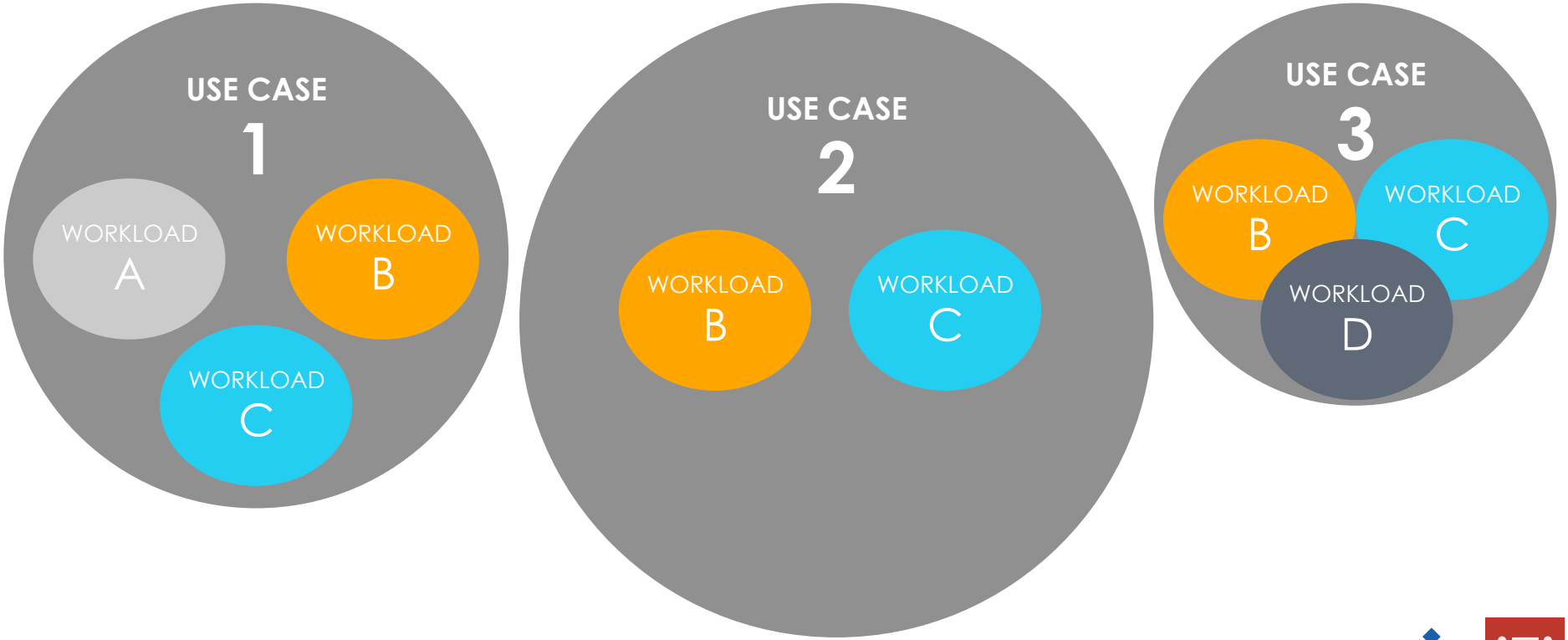
DATA STRATEGY

and choosing projects for maximum impact



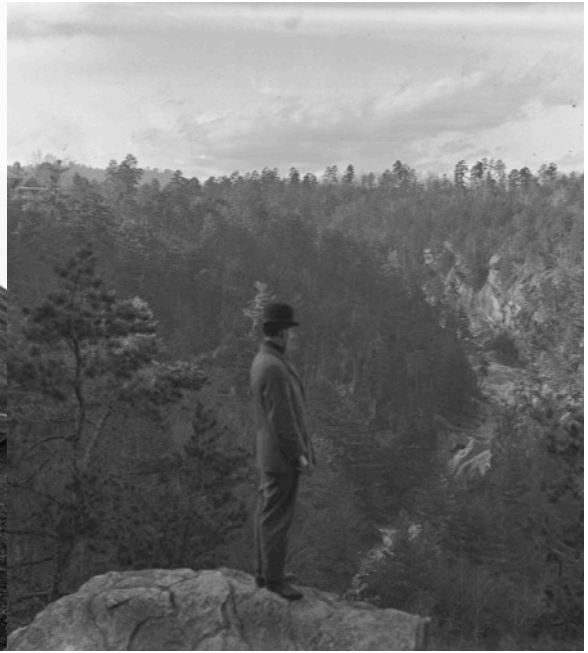
DEFINE YOUR ROADMAP

IDENTIFY STRATEGIC WORKLOADS



PRIORITIES

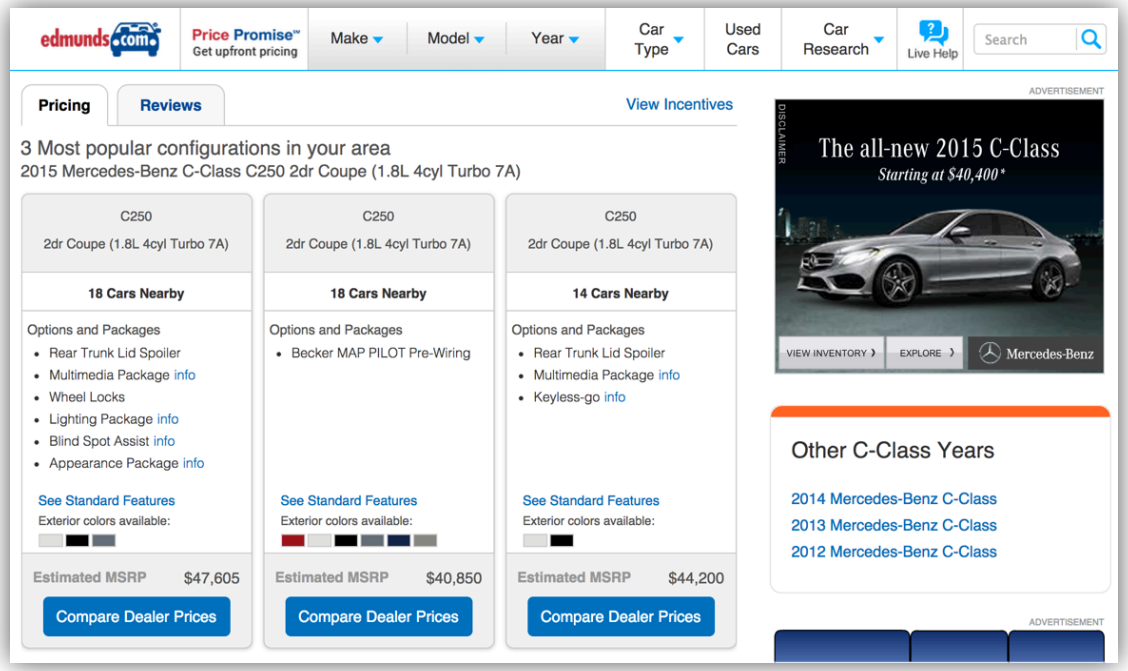
DIMENSIONS



ASSUMPTIONS
(overcome them)

LATHER, RINSE, REPEAT


Putting this into action for
EDMUNDS.com

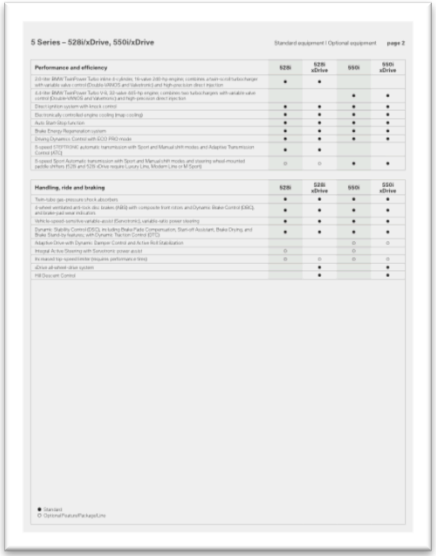


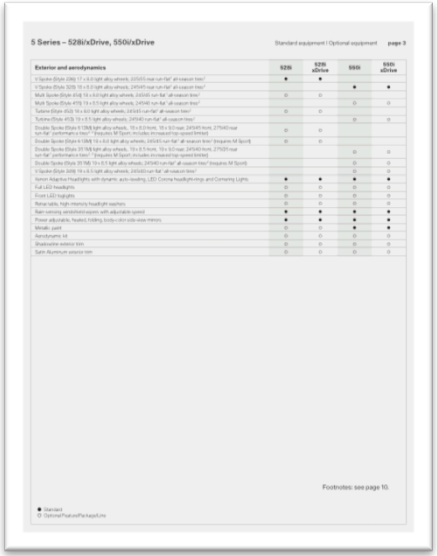
Existing revenue streams:

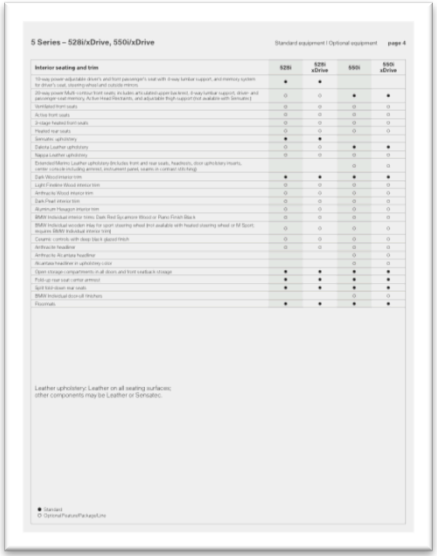
- Ads
- Price quotes (leads)

Shopping is the focus:

- Need real-time inventory
- Accurately described VIN's





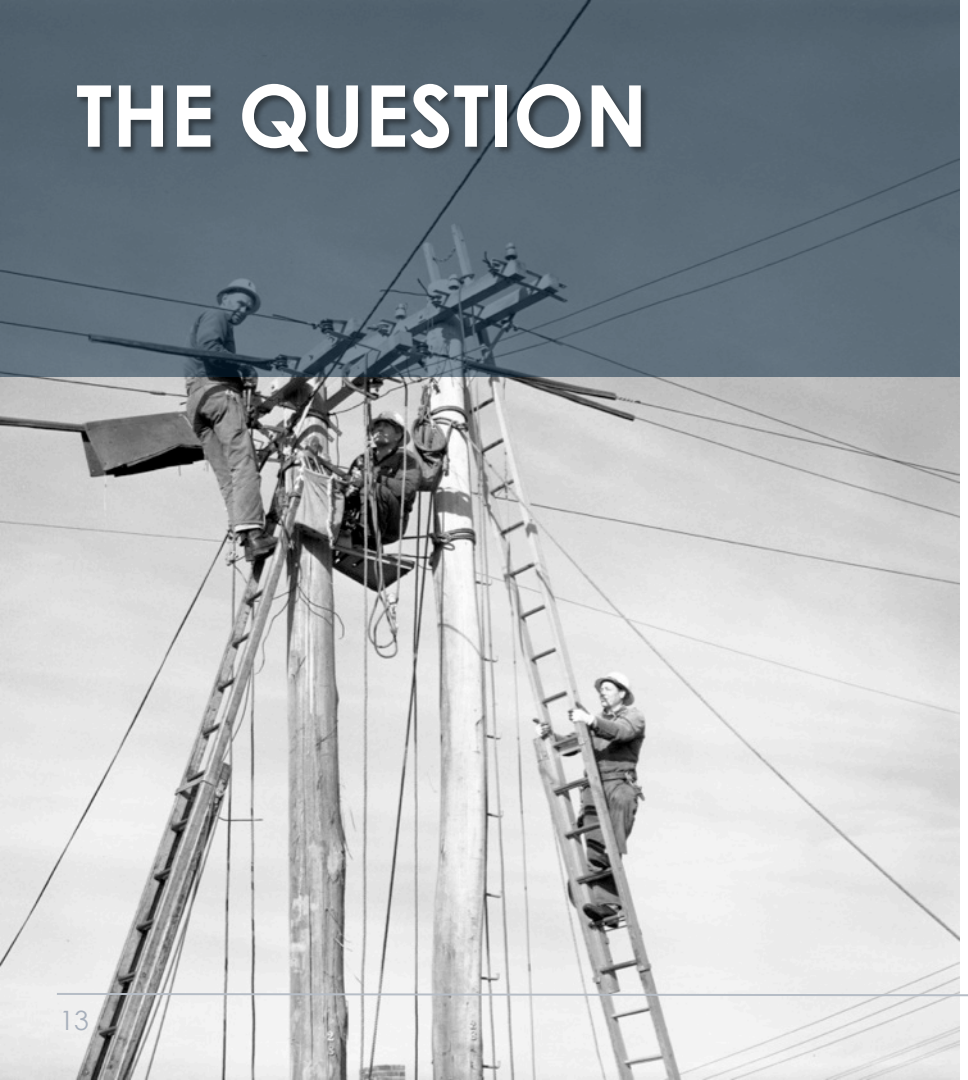


THE PDFs

THE CHALLENGE

- Couldn't keep pace with original equipment manufacturers (OEMs)
- Approach was largely manual, and backlogs would develop
- ~6.5% of VIN's being held back
- Content operations team was a silo

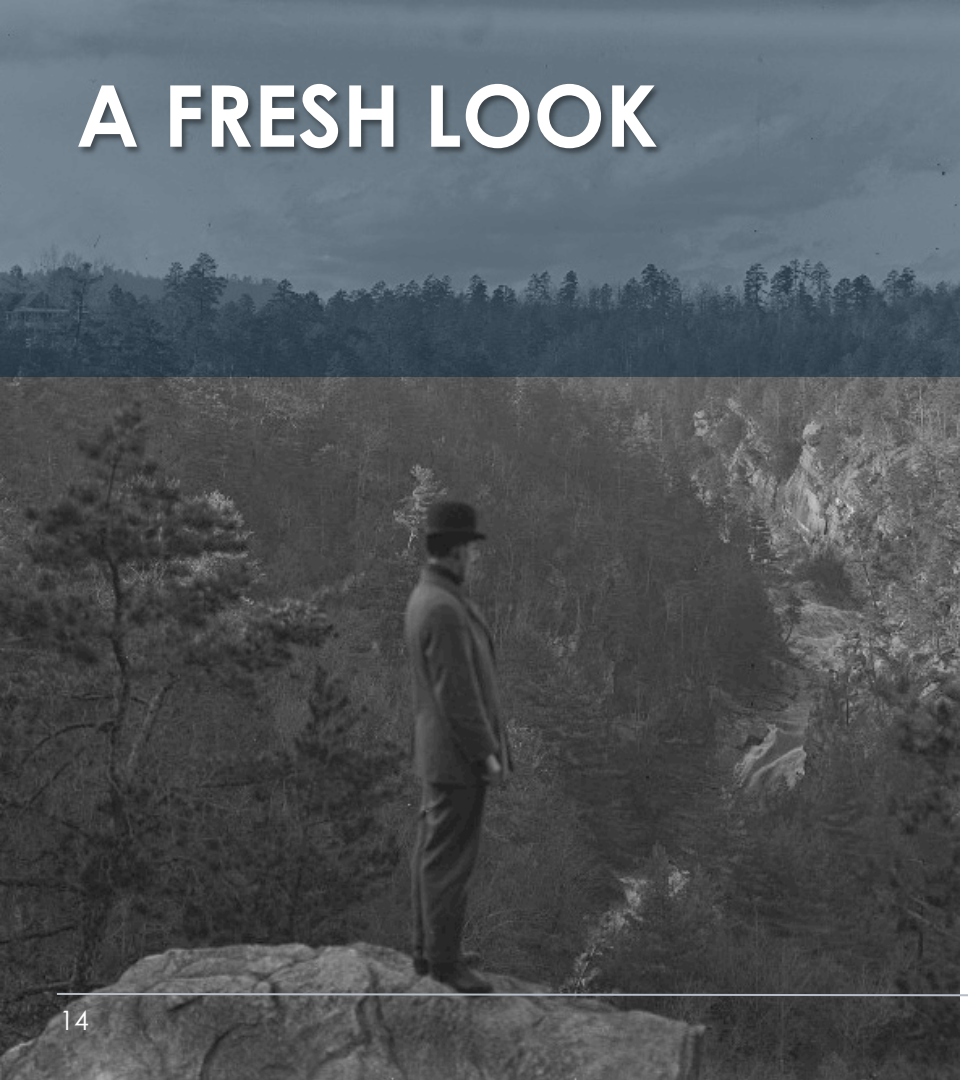
THE QUESTION



What is the
minimum viable data
required to get a VIN live?

(Then come back to add features and specs.)

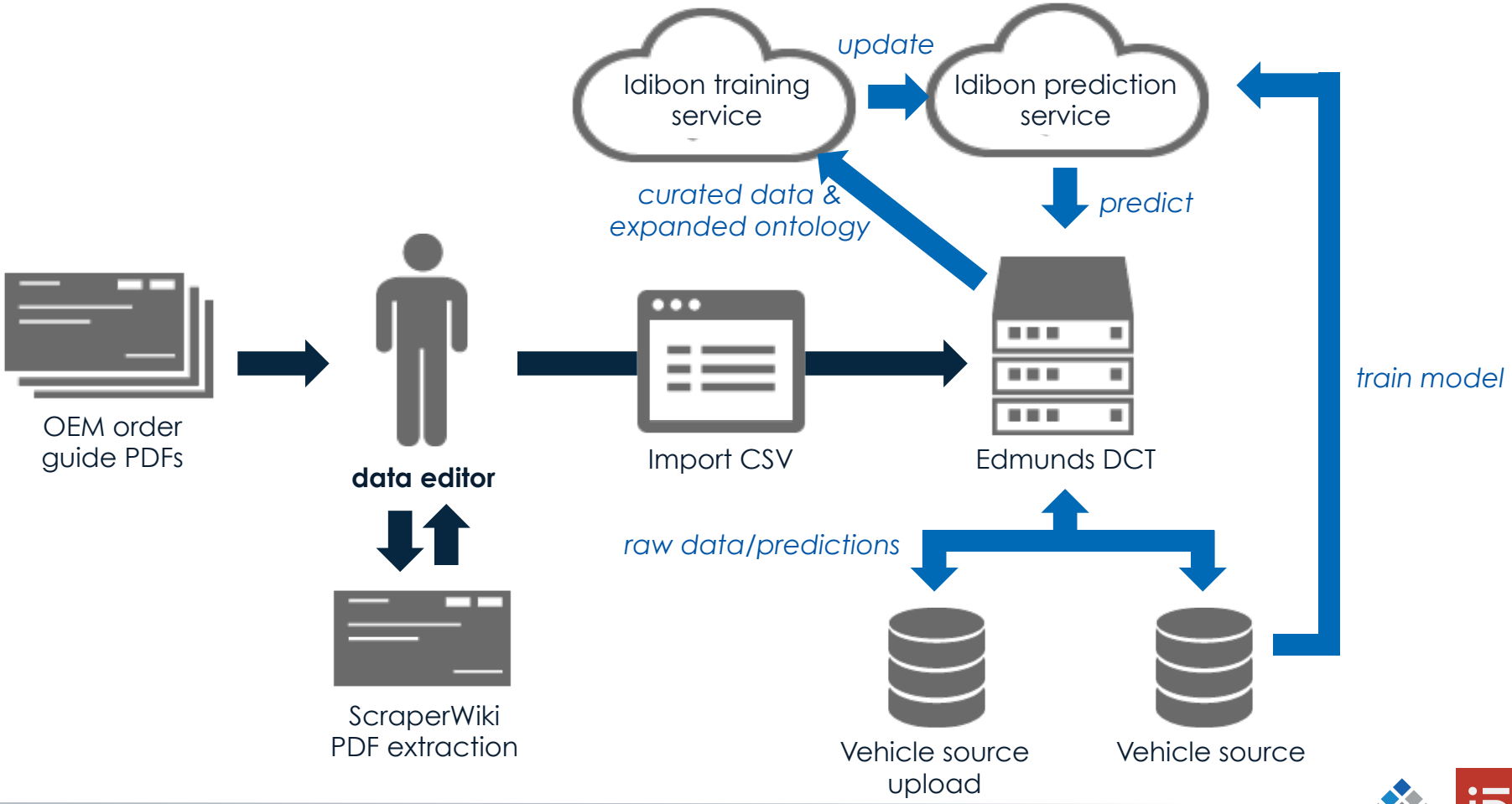
A FRESH LOOK

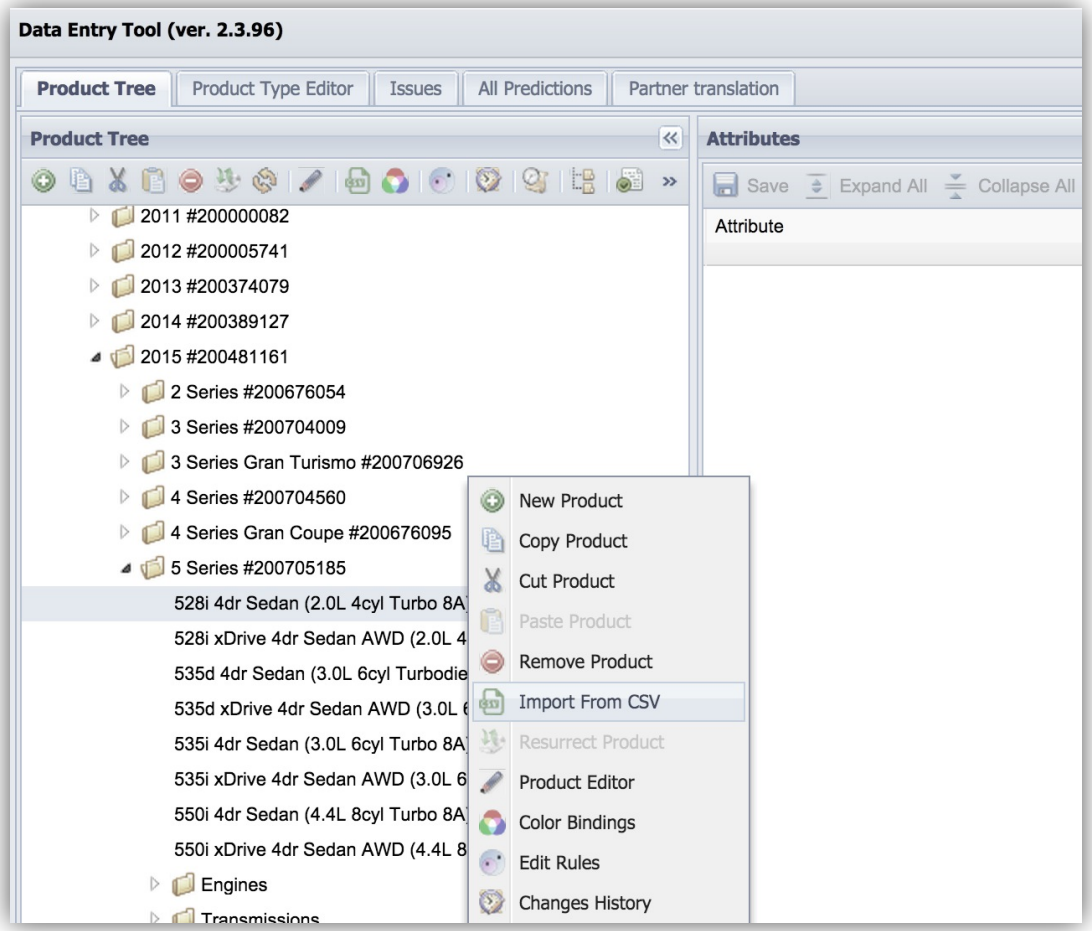


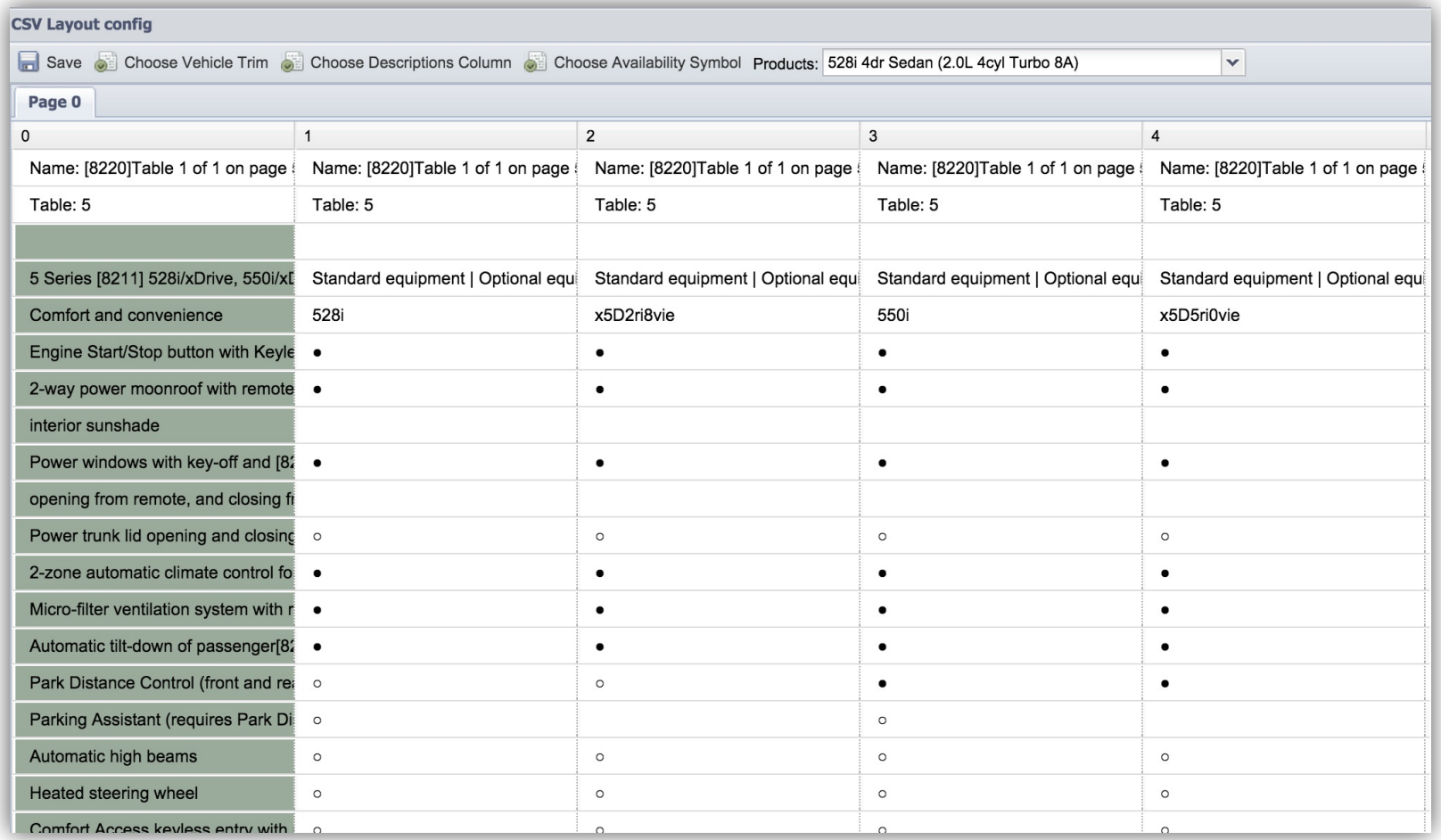
- Worked with Product Owners to define shell products (aka *minimally viable data*)
- Leveraged NLP to automate OEM data translations
- Fully integrated NLP workflow into existing tooling

Capabilities ***have*** to be integrated into the business process.

Ultimately, you may be able to make predictions, but putting those into action is the real thing.







CSV Layout config

Save | Choose Vehicle Trim | Choose Descriptions Column | Choose Availability Symbol | Products: 528i 4dr Sedan (2.0L 4cyl Turbo 8A)

Page 0

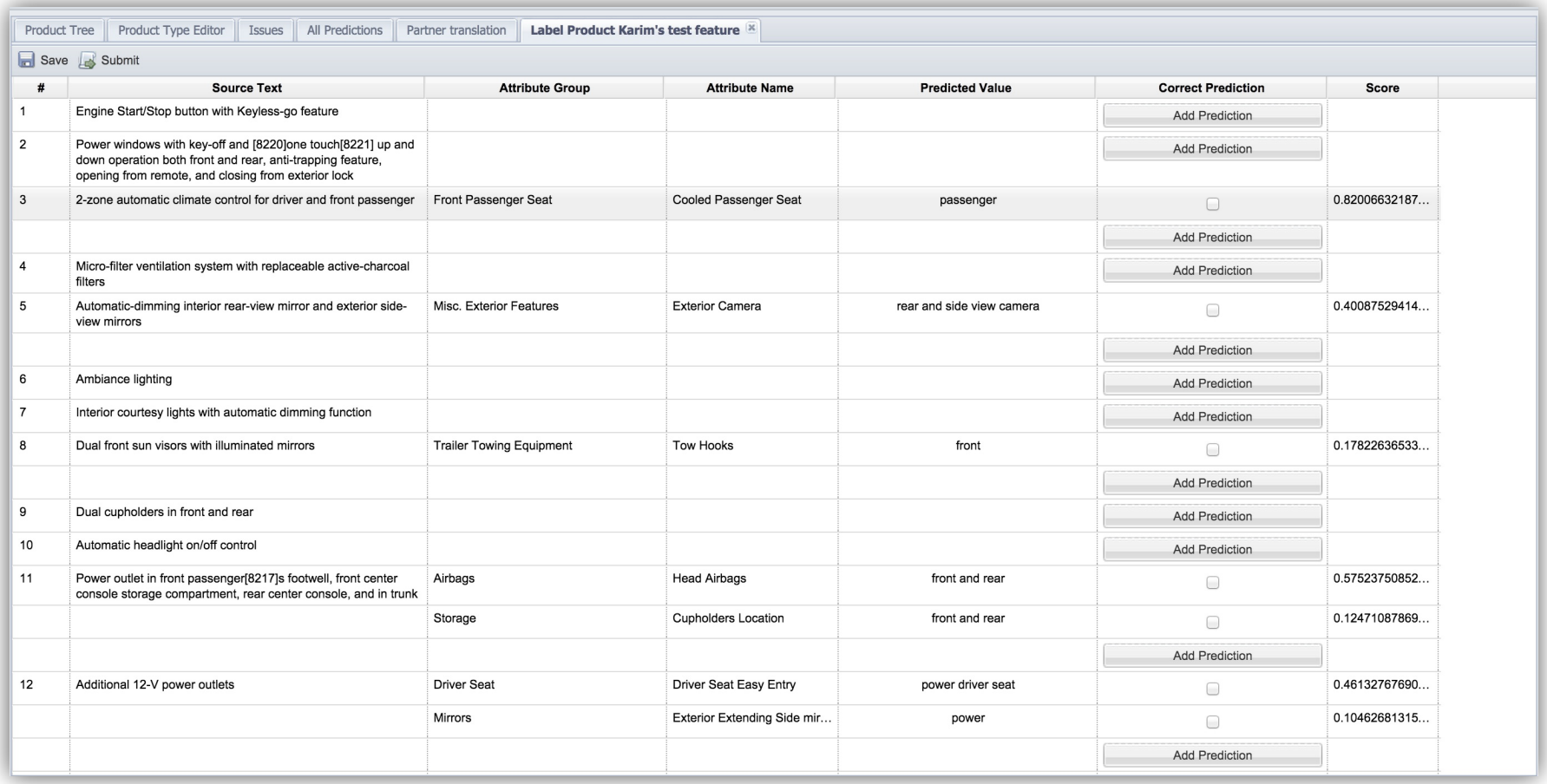
0	1	2	3	4
Name: [8220]Table 1 of 1 on page	Name: [8220]Table 1 of 1 on page	Name: [8220]Table 1 of 1 on page	Name: [8220]Table 1 of 1 on page	Name: [8220]Table 1 of 1 on page
Table: 5	Table: 5	Table: 5	Table: 5	Table: 5
5 Series [8211] 528i/xDrive, 550i/xD	Standard equipment | Optional equ	Standard equipment | Optional equ	Standard equipment | Optional equ	Standard equipment | Optional equ
Comfort and convenience	528i	x5D2ri8vie	550i	x5D5ri0vie
Engine Start/Stop button with Keyle	●	●	●	●
2-way power moonroof with remote	●	●	●	●
interior sunshade				
Power windows with key-off and [82	●	●	●	●
opening from remote, and closing fr				
Power trunk lid opening and closing	○	○	○	○
2-zone automatic climate control fo	●	●	●	●
Micro-filter ventilation system with r	●	●	●	●
Automatic tilt-down of passenger[82	●	●	●	●
Park Distance Control (front and rea	○	○	●	●
Parking Assistant (requires Park Di	○		○	
Automatic high beams	○	○	○	○
Heated steering wheel	○	○	○	○
Comfort Access keyless entry with	○	○	○	○

CSV Layout config

Save | Choose Vehicle Trim | Choose Descriptions Column | Choose Availability Symbol | Products: Karim's test feature

Page 0

0	1	2	3	4
Name: [8220]Table 1 of 1 on page	Name: [8220]Table 1 of 1 on page	Name: [8220]Table 1 of 1 on page	Name: [8220]Table 1 of 1 on page	Name: [8220]Table 1 of 1 on page
Table: 5	Table: 5	Table: 5	Table: 5	Table: 5
5 Series [8211] 528i/xDrive, 550i/xD	Standard equipment | Optional equ	Standard equipment | Optional equ	Standard equipment | Optional equ	Standard equipment | Optional equ
Comfort and convenience	528i	x5D2ri8vie	550i	x5D5ri0vie
Engine Start/Stop button with Keyle	•	•	•	•
2-way power moonroof with remote	•	•	•	•
interior sunshade				
Power windows with key-off and [82	•	•	•	•
opening from remote, and closing fr				
Power trunk lid opening and closing	○	○	○	○
2-zone automatic climate control fo	•	•	•	•
Micro-filter ventilation system with r	•	•	•	•
Automatic tilt-down of passenger[82	•	•	•	•
Park Distance Control (front and rea	○	○	•	•
Parking Assistant (requires Park Di	○		○	
Automatic high beams	○	○	○	○



Product Tree | Product Type Editor | Issues | All Predictions | Partner translation | **Label Product Karim's test feature**

Save | Submit

#	Source Text	Attribute Group	Attribute Name	Predicted Value	Correct Prediction	Score
1	Engine Start/Stop button with Keyless-go feature				Add Prediction	
2	Power windows with key-off and [8220]one touch[8221] up and down operation both front and rear, anti-trapping feature, opening from remote, and closing from exterior lock				Add Prediction	
3	2-zone automatic climate control for driver and front passenger	Front Passenger Seat	Cooled Passenger Seat	passenger	☐	0.82006632187…
					Add Prediction	
4	Micro-filter ventilation system with replaceable active-charcoal filters				Add Prediction	
5	Automatic-dimming interior rear-view mirror and exterior side-view mirrors	Misc. Exterior Features	Exterior Camera	rear and side view camera	☐	0.40087529414…
					Add Prediction	
6	Ambiance lighting				Add Prediction	
7	Interior courtesy lights with automatic dimming function				Add Prediction	
8	Dual front sun visors with illuminated mirrors	Trailer Towing Equipment	Tow Hooks	front	☐	0.17822636533…
					Add Prediction	
9	Dual cupholders in front and rear				Add Prediction	
10	Automatic headlight on/off control				Add Prediction	
11	Power outlet in front passenger[8217]s footwell, front center console storage compartment, rear center console, and in trunk	Airbags	Head Airbags	front and rear	☐	0.57523750852…
		Storage	Cupholders Location	front and rear	☐	0.12471087869…
					Add Prediction	
12	Additional 12-V power outlets	Driver Seat	Driver Seat Easy Entry	power driver seat	☐	0.46132767690…
		Mirrors	Exterior Extending Side mir…	power	☐	0.10462681315…
					Add Prediction	

Hierarchical Entity Resolution

'Micro-filter ventilation system with replaceable active-charcoal filters'

Attribute Group

Attribute Name

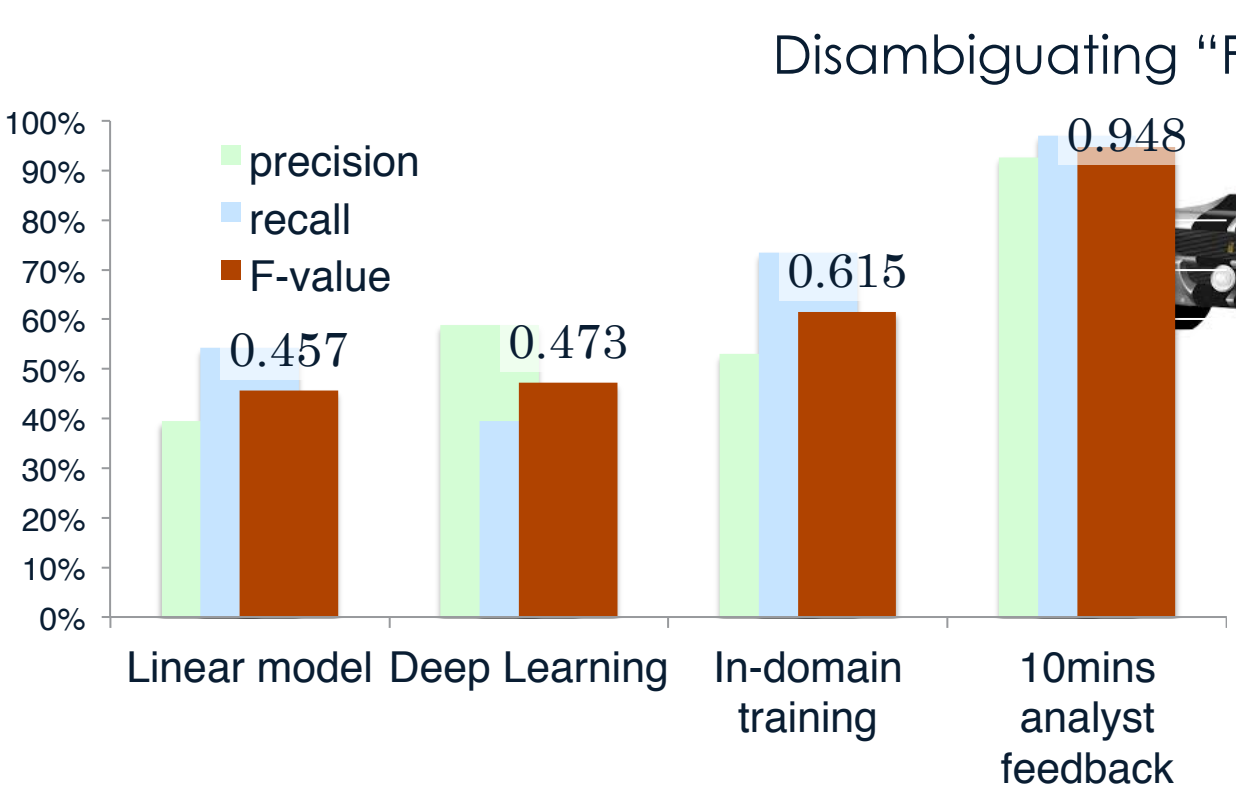
Attribute Value

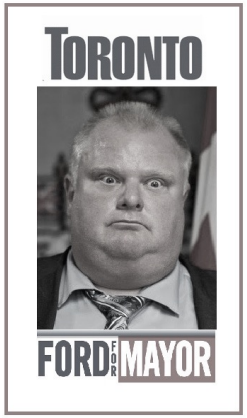
Air_Conditioning

Air_Filtration

Interior_Active_
Charcoal_Filter

Data beats algorithms; feedback beats data



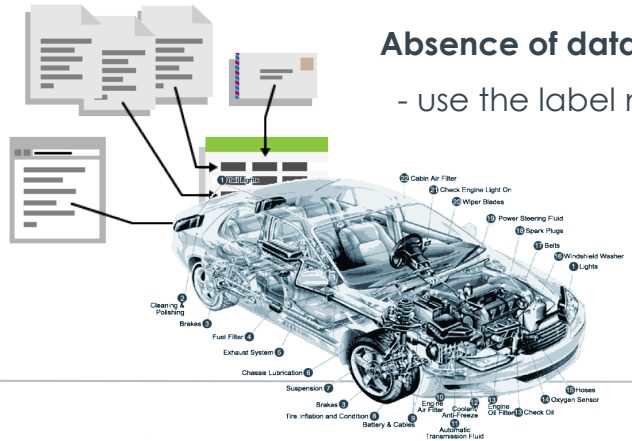


Edmunds ontology

'Micro-filter ventilation system with replaceable active-charcoal filters'

Air_Conditioning (depth 1)

- Air_Filtration (depth 2)
  - Interior_Active_Charcoal_Filter (depth 3)
  - …
- Climate_Control_Memory
  - …
- Front_Air_Conditioning
  - Front Zone
- Rear_Air_Conditioning
  - Rear Zones
- …



4000 (!) car features

20,000 labeled items to date

- requires expertise; accuracy with few data points

Features:

- words, n-grams, word shapes, taxonomies

Negative sampling:

- other items within the same parent

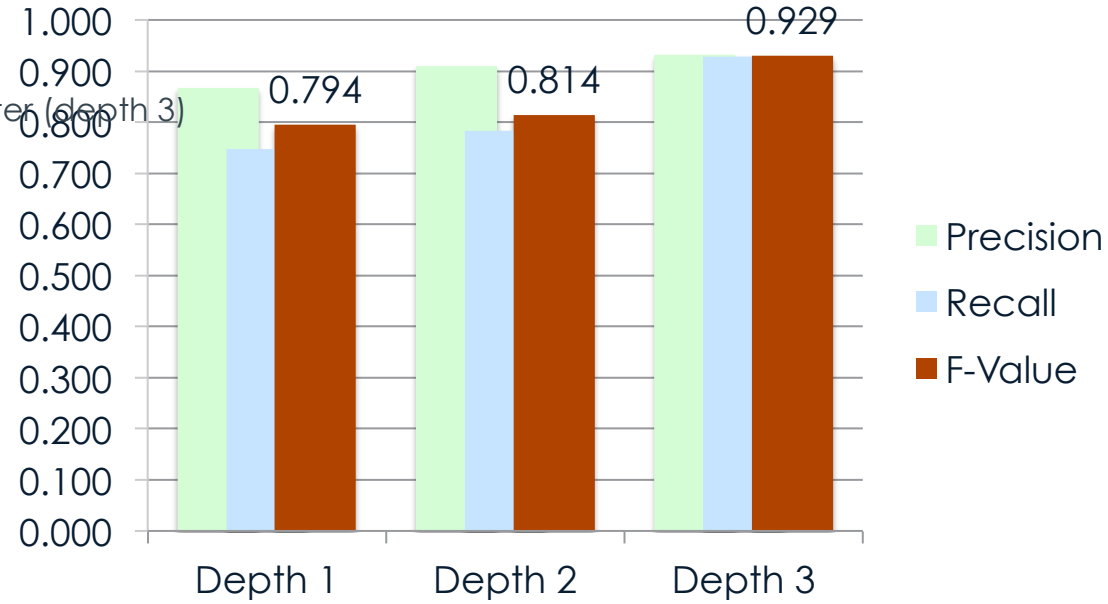
Absence of data:

- use the label name as the first data item

Accuracy at depths in the ontology

Air_Conditioning (depth 1)

- Air_Filtration (depth 2)
  - Interior_Active_Charcoal_Filter (depth 3)
  - …
- Climate_Control_Memory
  - …
- Front_Air_Conditioning
  - Front Zone
- Rear_Air_Conditioning
  - Rear Zones
- …



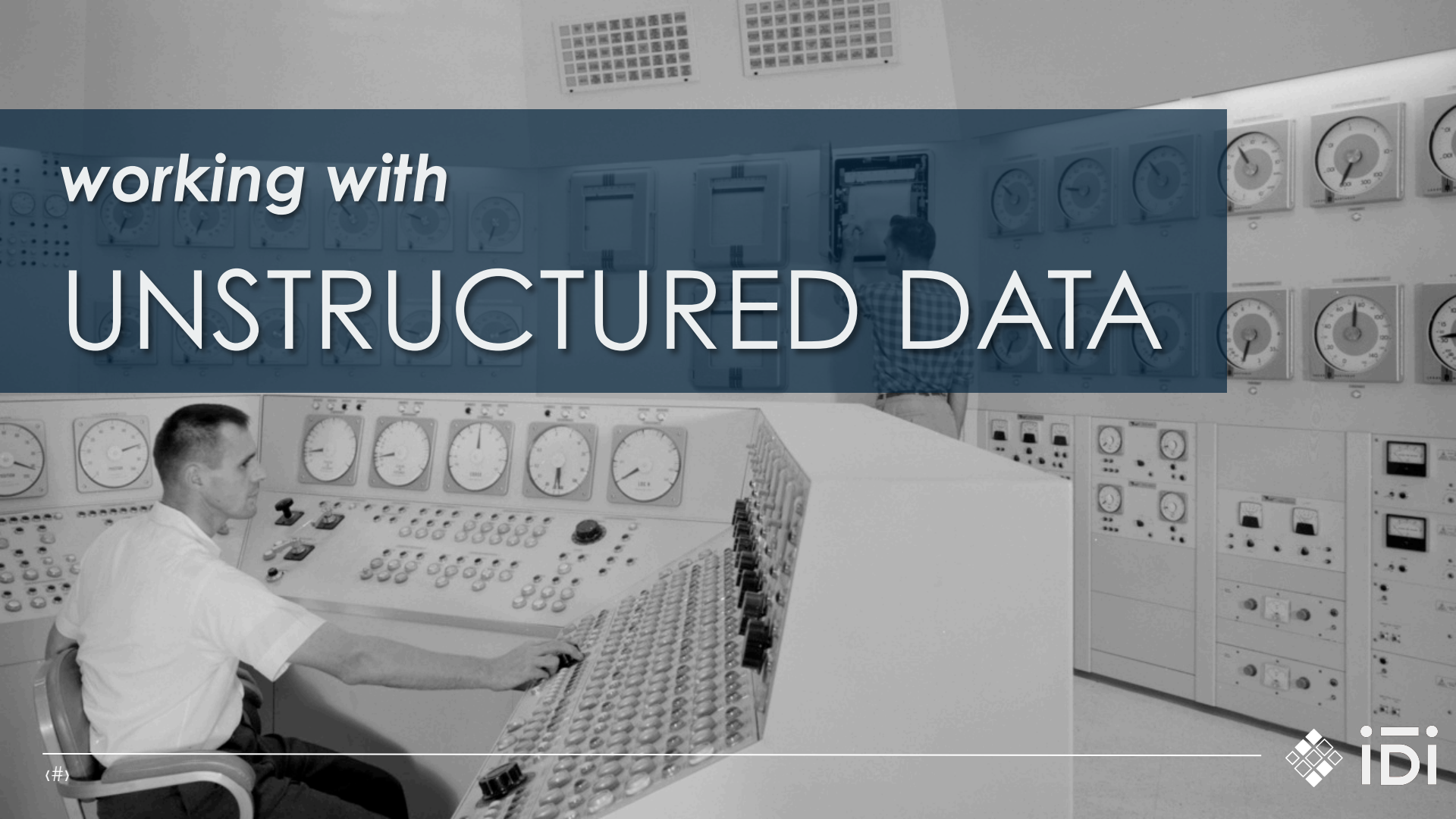
Controlled vocabulary allows accuracy with few data points

NEW CAPABILITIES

- 1–2 days to get a new Car Model live online vs. 2 weeks (~85% reduction)
- 95% reduction in backlog
- API for making predictions on unstructured vehicle data using Edmunds' ontology assets

TAKEAWAYS

- Use data to highlight the real issues
- Change is difficult, agile methods help users embrace new approaches
- New processes must adapt to workflows



working with

UNSTRUCTURED DATA

Entity Resolution

Recognizing references to objects: "BMW 435" = 2014 BMW 4 Series 435i xDrive 2dr Coupe AWD (3.0L 6cyl Turbo 8A)

Concept Discovery

AWD = 4WD = xDrive = quattro
Manual transmission = stick shift

Sentiment Analysis

Positive comment referred to exterior design of competitor's model

What Others Are Saying

Consumer Reviews for the BMW 4 Series

★★★★½ See all 7 reviews

1 of 1 people found this review helpful
★★★★½
Try the 428i
by **professorx** on Sep 3, 2014
Vehicle: 2014 BMW 4 Series 428i SULEV 2dr Coupe (2.0L 4cyl Turbo 8A)

Picked up my 428i yesterday. Was worried that the 4 cyl wouldn't be enough, and that I would not be

8 of 22 people found this review helpful
★★★★½
I (almost) love everything about
by **jrbskissummit** on Apr 13, 2014
Vehicle: 2014 BMW 4 Series 435i xDrive 2dr Coupe AWD (3.0L 6cyl Turbo 8A)

As far as I could tell there were only two cars that met my desires for a coupe with AWD, Manual transmission and forced induction for Colorado altitudes. The BMW 435 ended up being my choice. The other choice was the A5/S5.The 428 would have been a competitor for the A5 if it was available with manual. Anyway I drove a S5 and compared it with the 435. I preferred the S5 looks but engine dynamics are noticeably better in the 435 (if the sport mode is active). My car has a manual, so sport mode doesn't fool with the transmission shift points as it does in the Automatic cars. As expensive as it is the 435 seems a better value than the S5. I wish it was all wonderful but it could be better.

Want these slides? Go to:

TO ADD

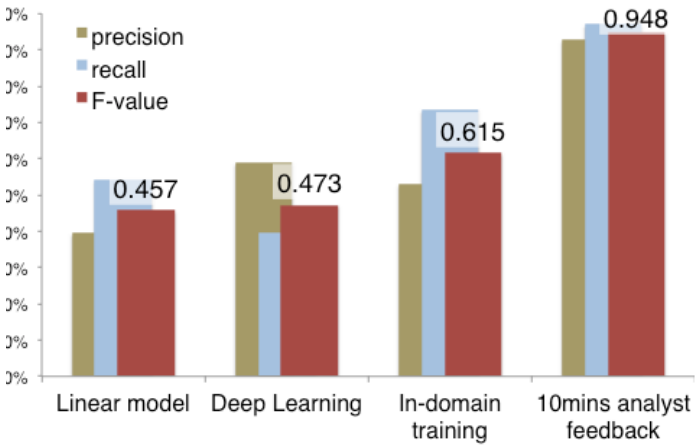
SILICON VALLEY DATA SCIENCE

IDIBON

How do we get the analysts' feedback?

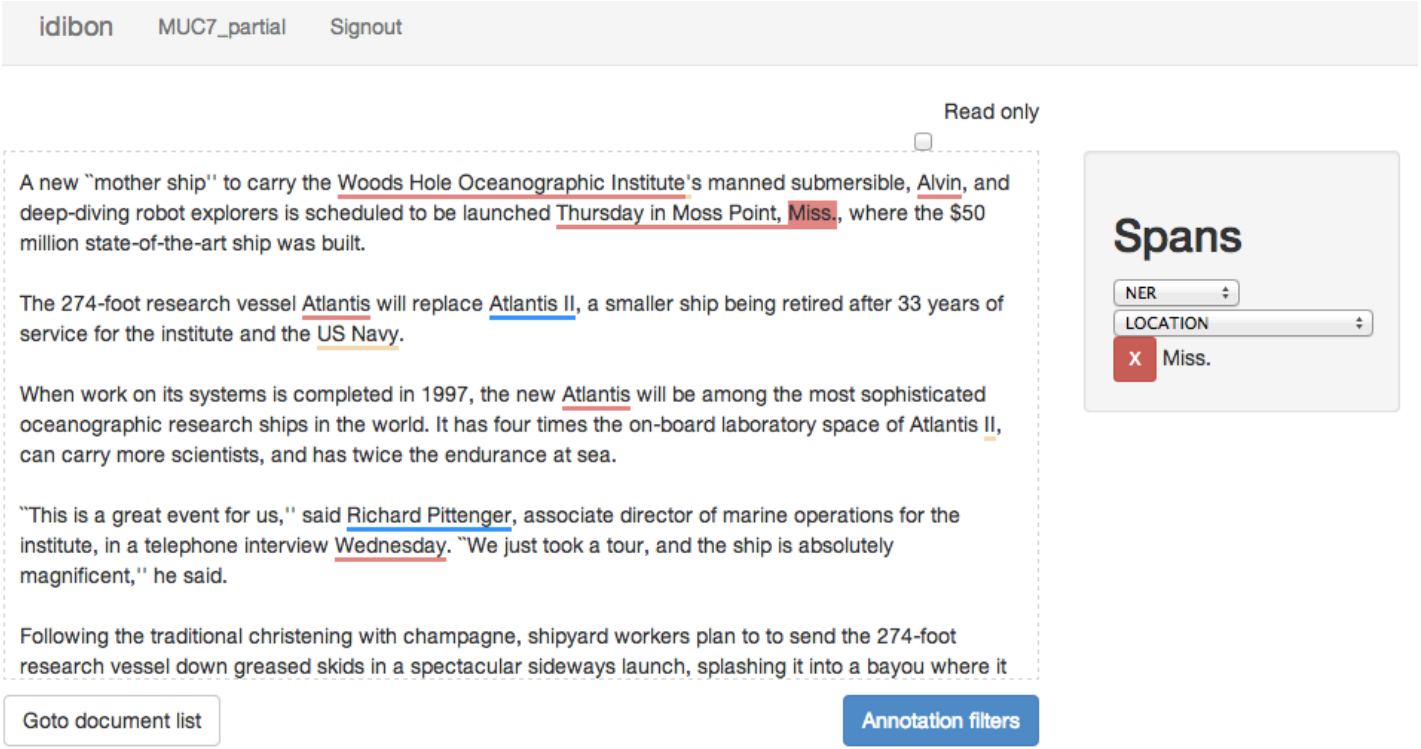
How many ways could we *ask* someone to distinguish the right "Ford"?

Get it right, and save 90% of the labor ≈ 90% of total cost.

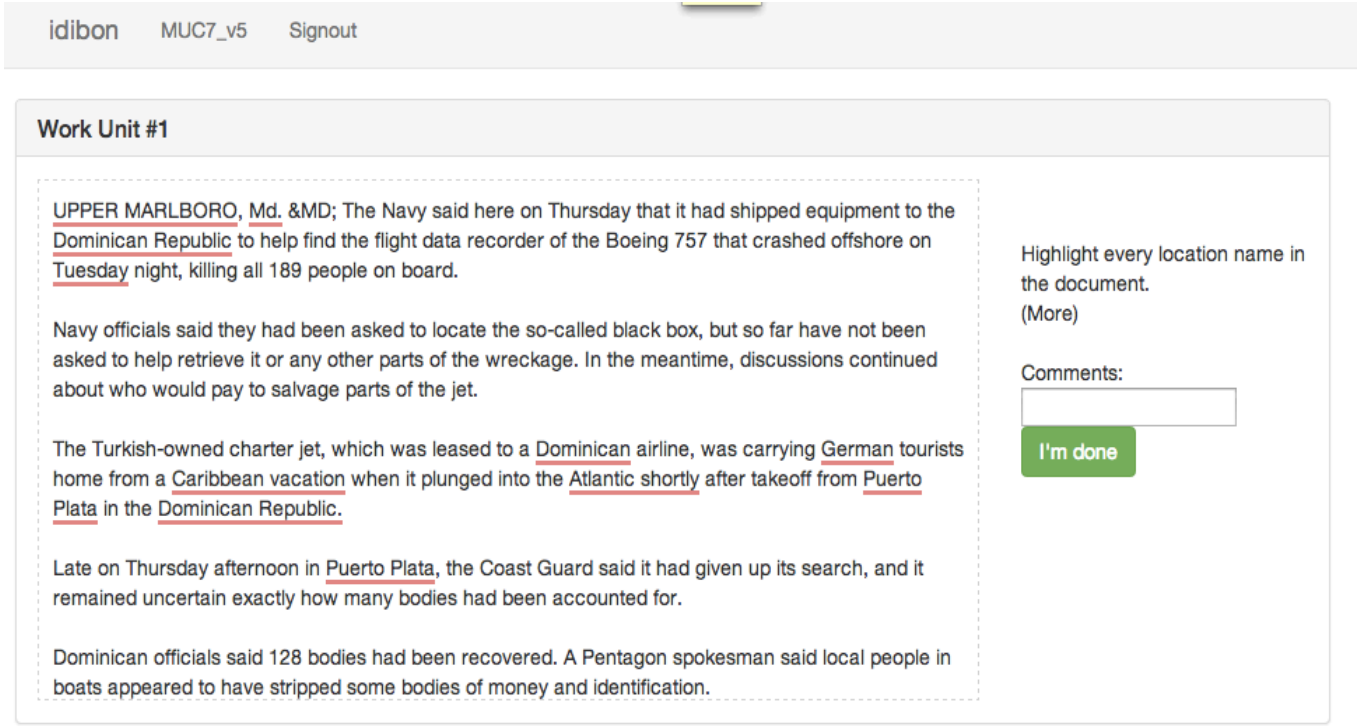


Distinguishing "Ford" the company from people called "Ford"

Super-user, guided by model suggestions



Targeting one label at a time



Simple accept/reject annotation for fast annotation