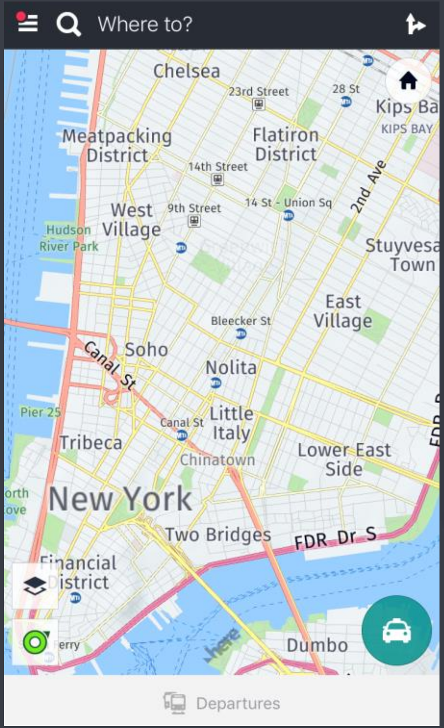
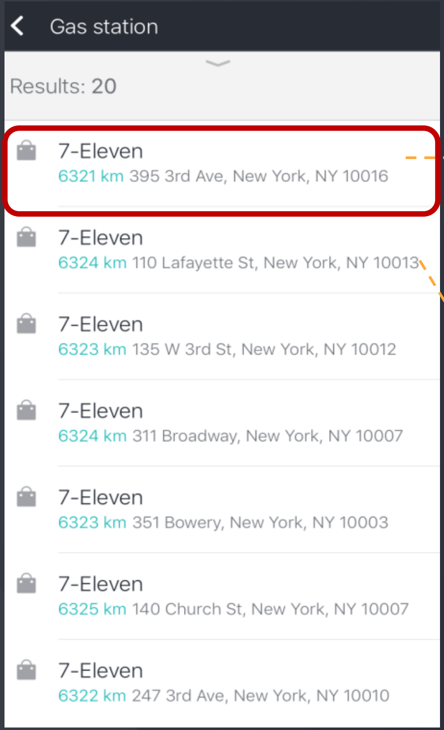
Category or Cate…strophe?

Soojung Hong
Lead Research Engineer @ Zurich, Switzerland
May 22. 2019, HERE AI Summit

Search query *'Gas station'*





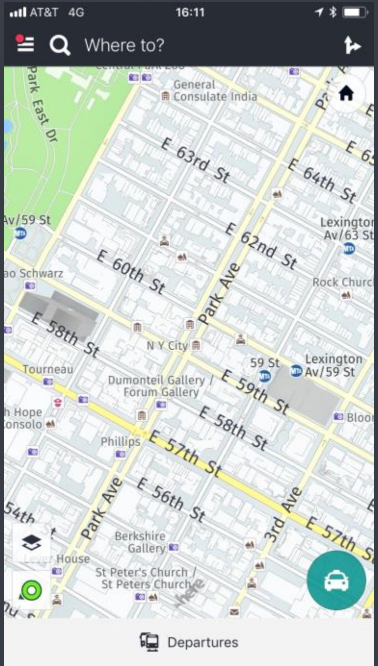
395 3rd Ave, New York

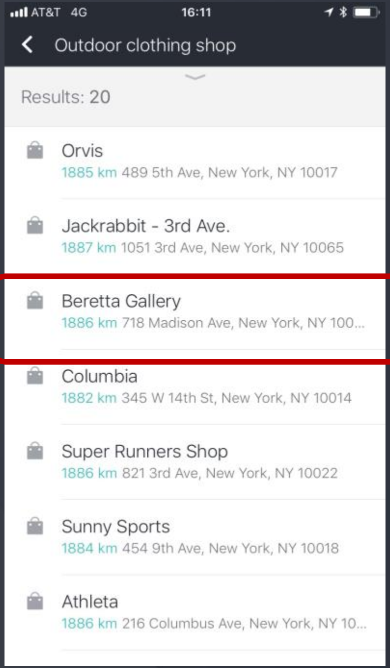
110 Lafayette St, New York

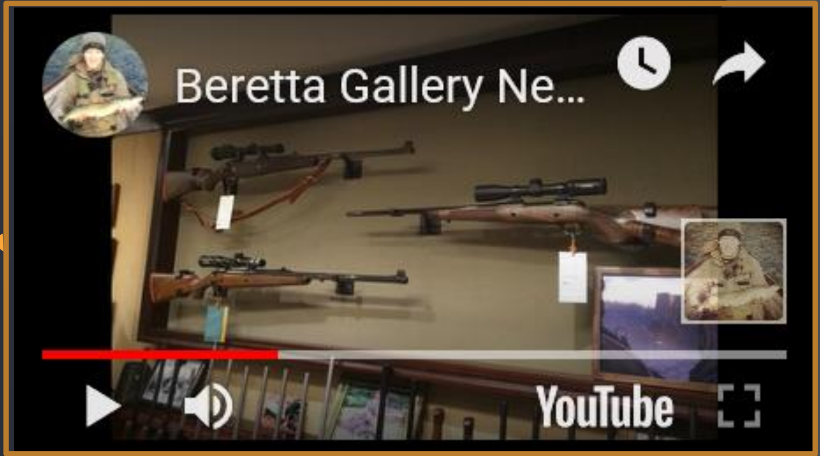
Ground truth

Viewport : *'New York City'*

Query ‘*Outdoor clothing shop*’







https://www.youtube.com/watch?v=eL-9K9rPVUo

Ground truth

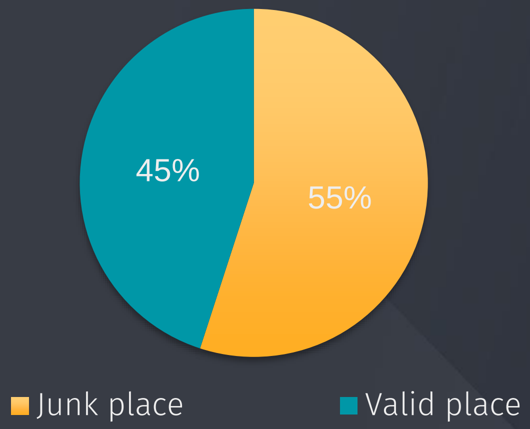
Viewport : ‘*New York City*’

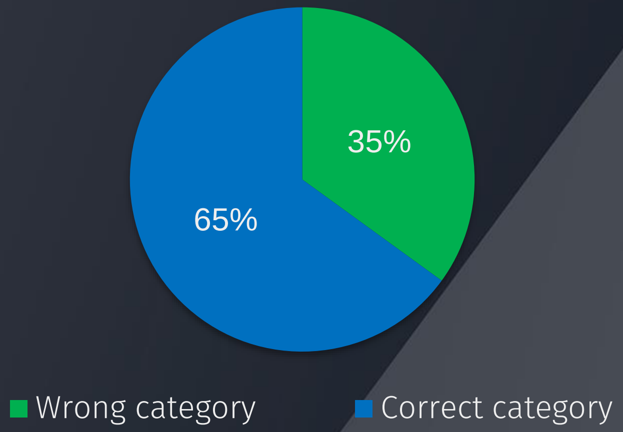
here

Are these only a few anecdotes?

According to Place Quality Auditing Team

- TQS (Tiered Quality Scoring) on sample places by field engineers
- 230M places registered in PDS (Place Data Store)
- All places are mapped with at least one of 440 HERE Categories





Customer complaint is a best way to fix?

SoMo

HERE WeGo

HERE Search/Recommendation

Location Graph

HERE internal

Place (Category)

GARMIN

BMW

HYUNDAI

amazon

Customers of HERE Place

HERE Categories Now

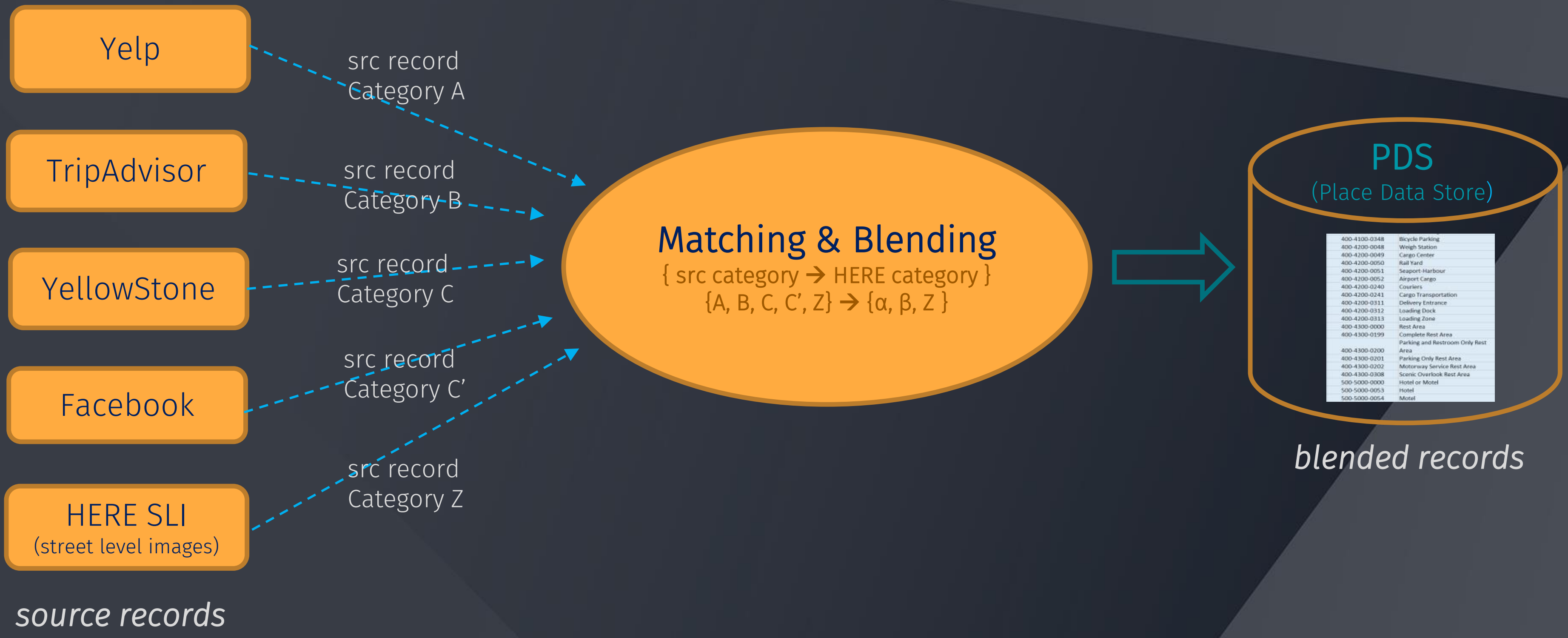
A perfect problem to be solved by ML

- Large scale data : 230M + α places in the world
- Hard to verify all places by field engineers manually
- Need to *minimize* subjective interpretation about places
- Detect noisy information automatically

Place Name	Categories	Food Types
7-Eleven	[" 'Convenience Store'", " 'Restaurant'", " 'ATM'"]	[]
7-Eleven	[" 'Petrol-Gasoline Station'", " 'Convenience Store'", " 'Restaurant'"]	[]
7-Eleven	[" 'Convenience Store'", " 'Deli'", " 'Restaurant'"]	[" 'Sandwich'"]
7-Eleven	[" 'ATM'", " 'Convenience Store'", " 'Restaurant'"]	[]
7-Eleven	[" 'Convenience Store'", " 'Restaurant'", " 'ATM'"]	[]
7-Eleven	[" 'Specialty Store'", " 'Convenience Store'", " 'Fast Food'", " 'Restaurant'"]	[" 'Pizza'"]
7-Eleven	[" 'Convenience Store'", " 'Restaurant'"]	[" 'Pizza'", " 'Sandwich'"]
7-Eleven ✓ ***incorrect***	[" 'Petrol-Gasoline Station'", " 'Convenience Store'", " 'Restaurant'"]	[]
7-Eleven	[" 'Convenience Store'", " 'Restaurant'", " 'ATM'"]	[]
7-Eleven	[" 'Convenience Store'", " 'Restaurant'"]	[]
7-Eleven	[" 'Convenience Store'", " 'Casual Dining'", " 'Restaurant'", " 'ATM'"]	[" 'International'"]
7-Eleven	[" 'Convenience Store'", " 'Restaurant'"]	[]
7-Eleven	[" 'Convenience Store'", " 'Restaurant'"]	[]
7-Eleven	[" 'Petrol-Gasoline Station'", " 'Convenience Store'", " 'Restaurant'"]	[]
New Rebozo Chicago	[" 'Restaurant'"]	[" 'Mexican'"]
Buzz Bait Taqueria	[" 'Fast Food'", " 'Taqueria'", " 'Food Market-Stall'", " 'Restaurant'"]	[" 'Mexican'", " 'Seafood'"]
Velvet Taco Chicago	[" 'Fast Food'", " 'Restaurant'"]	[" 'American'", " 'Mexican'"]
Seoul Taco	[" 'Casual Dining'", " 'Restaurant'"] ✓ ***redundant***	[" 'Mexican'", " 'Asian'", " 'Korean'"]
Taco Joint River North	[" 'Casual Dining'", " 'Fast Food'", " 'Taqueria'", " 'Restaurant'"]	[" 'Mexican'"]
Chipotle	[" 'Fast Food'", " 'Restaurant'"]	[" 'Mexican'"]
Adobo Grill Old Town	[" 'Casual Dining'", " 'Restaurant'", " 'Bar or Pub'"]	[" 'American-Southwestern'", " 'Mexican'", " 'Vegetarian'", " 'Grill'"]
North & Clark CafÃ©	[" 'Casual Dining'", " 'Restaurant'", " 'Coffee Shop'"]	[" 'American'"]
Adobe Gila's	[" 'Restaurant'"]	[" 'American'", " 'American-Southwestern'", " 'Mexican'"]
Foodlife	[" 'Casual Dining'", " 'Restaurant'"]	[" 'American'", " 'Mexican'", " 'Chinese'", " 'Pizza'", " 'International'"]
Downtown Dogs	[" 'Fast Food'", " 'Restaurant'"]	[" 'American'", " 'Mexican'", " 'Hot Dogs'"]
Blue Agave Tequila Bar & Restaurant	[" 'Casual Dining'", " 'Restaurant'", " 'Bar or Pub'"]	[" 'American'", " 'American-Southwestern'", " 'Mexican'"]
Sedgwick's Bar & Grill	[" 'Casual Dining'", " 'Restaurant'", " 'Bar or Pub'"]	[" 'American'", " 'American-Southwestern'", " 'Mexican'", " 'Grill'"]
Zia	[" 'Restaurant'"]	[" 'Mexican'"]
Dublin's Bar & Grill	[" 'Casual Dining'", " 'Nightlife-Entertainment'", " 'Restaurant'", " 'Bar or Pub'"]	[" 'Mexican'", " 'Irish'", " 'Grill'"]
Salpicon	[" 'Fine Dining'", " 'Restaurant'"]	[" 'Mexican'"]
Flaco's Tacos	[" 'Casual Dining'", " 'Restaurant'"]	[" 'Mexican'", " 'Chicken'"]
Nacional 27	[" 'Casual Dining'", " 'Dancing'", " 'Restaurant'"]	[" 'Mexican'", " 'Cuban'", " 'Spanish-Tapas'", " 'South American'", " 'Latin American'", " 'Internationa
Tallboy Taco	[" 'Fast Food'", " 'Restaurant'", " 'Bar or Pub'"]	[" 'Mexican'"]
Chipotle	[" 'Fast Food'", " 'Restaurant'"]	[" 'Mexican'"]
Su Casa Mexican Restaurant	[" 'Casual Dining'", " 'Restaurant'"]	[" 'American-Southwestern'", " 'Mexican'"]
Chipotle	[" 'Fast Food'", " 'Restaurant'"]	[" 'Mexican'"]
Patron's Hacienda	[" 'Casual Dining'", " 'Dancing'", " 'Restaurant'", " 'Bar or Pub'", " 'Take Out and Delivery Only'"]	[" 'Mexican'"]
Downtown Pizza	[" 'Fast Food'", " 'Restaurant'"]	[" 'Pizza'"]
Taco King	[" 'Restaurant'"] ✓ ***contradictory***	[" 'Mexican'"]
Burrito Bistro	[" 'Restaurant'"]	[" 'Mexican'"]

Example : place records of category 'restaurant (100-1000-0000)'

Place Pipeline & Noise



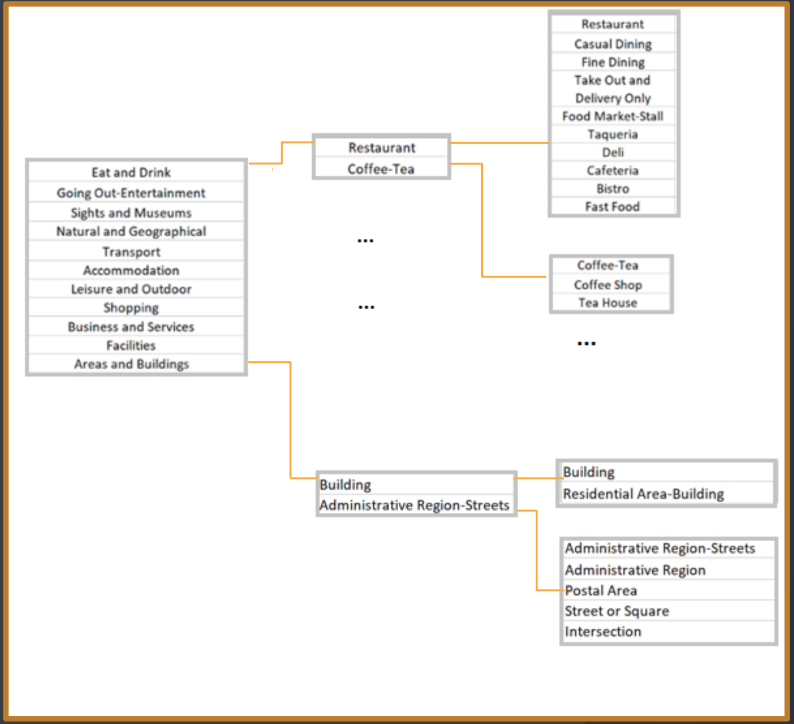
(Matching Example) Yelp : *'guns_and_ammo'*
Facebook : *'mens/women clothing', 'Accessary'*
85% of Facebook data is junk

PDS : *'sport goods', 'Men/Women Apparel'*

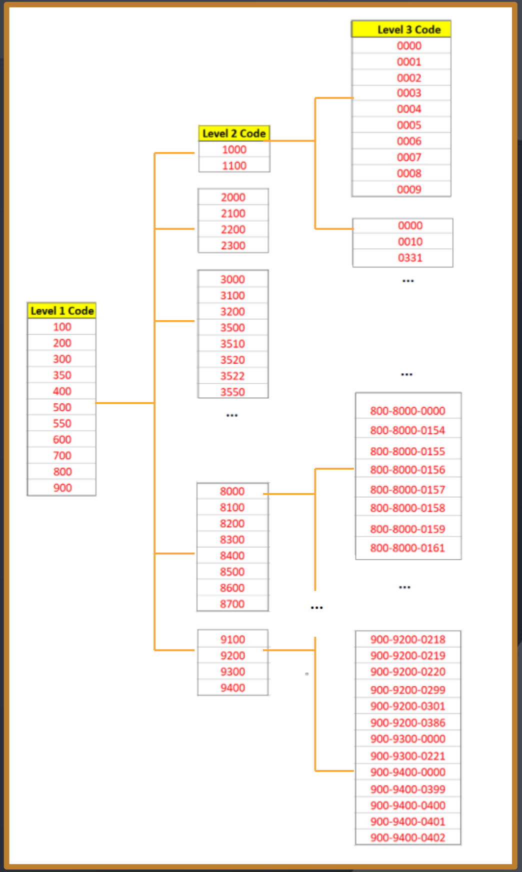
XXX-YYYY-ZZZZ

{Level 1 – Level 2 – Level 3}

Total : 440 Categories



Category Name

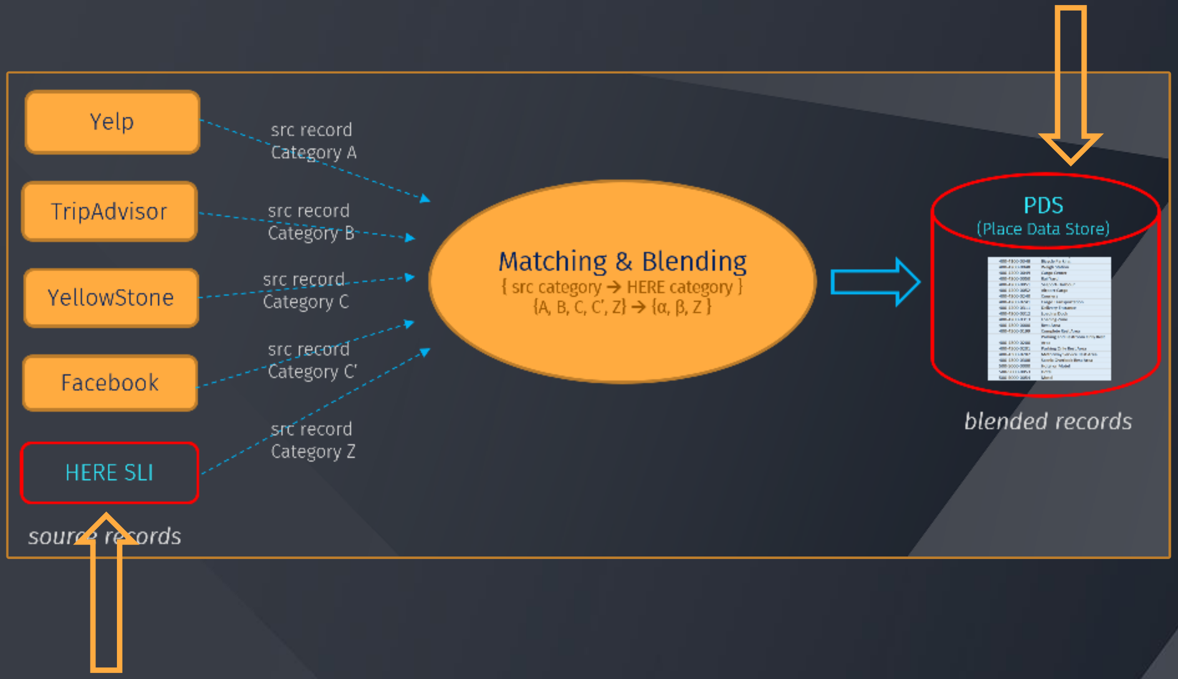


Category Code

here

How we tackle (Work In Progress)

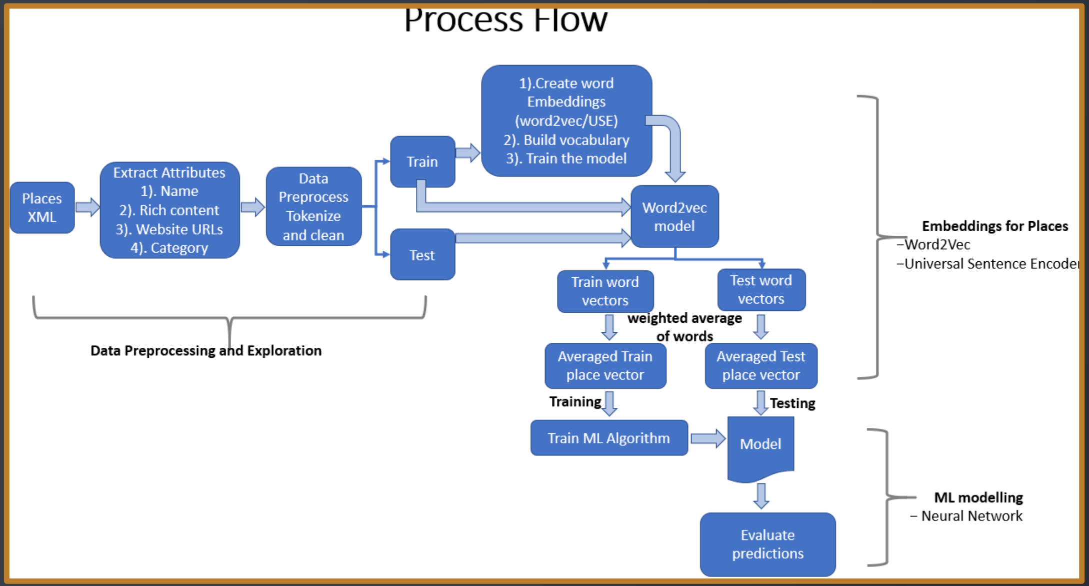
Category Relevance Model



Category Derivation Model

Category Derivation Model

Text Extraction From SLI → Attribute used: Place Name → NLP Embeddings For Input Text → Category Derivation Model Using Deep Learning



	precision	recall	f1-score
100	0.75	0.74	0.74
200	0.61	0.57	0.59
300	0.85	0.80	0.83
350	0.89	0.90	0.89
400	0.75	0.70	0.72
500	0.76	0.65	0.70
550	0.71	0.71	0.71
600	0.61	0.58	0.60
700	0.55	0.51	0.53
800	0.63	0.70	0.66
900	0.58	0.75	0.66
avg / total	0.69	0.68	0.69

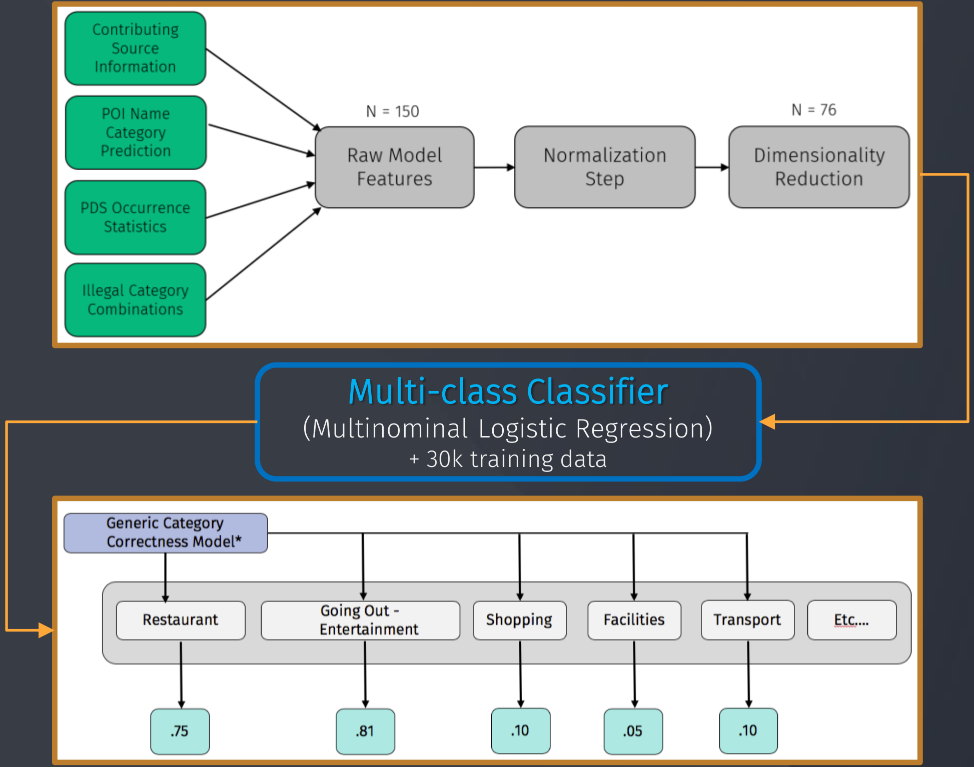
Level 1 : ~ 68% accuracy
Level 1 – Level 2 : 60%
Level 1 – Level 2 – Level 3 : ??

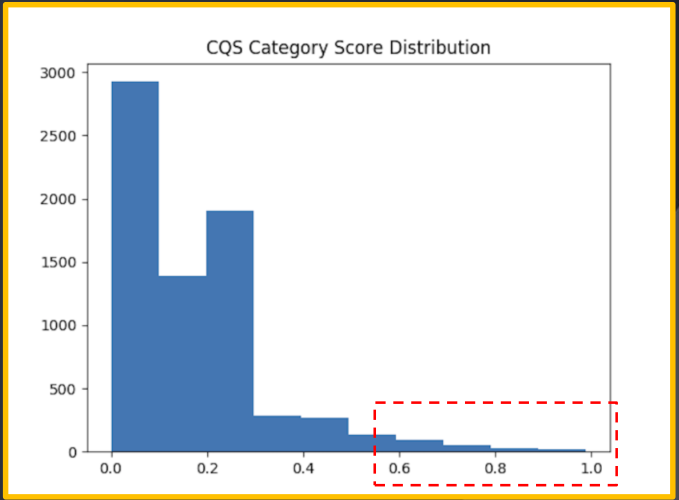
- Training data : 88,000 samples per category
- Test data : 15,000 samples per category
- Neural network with 3 hidden layers, *adam* optimizer, *softmax* output layer

Credit : Charitha H.

Category Relevance Model

- framework produces probabilistic assessments of category accuracy
- indicate *'primary'* category with highest relevance score





Test results
~ 5k USA place records

Credit : Matthew Tetkoskie

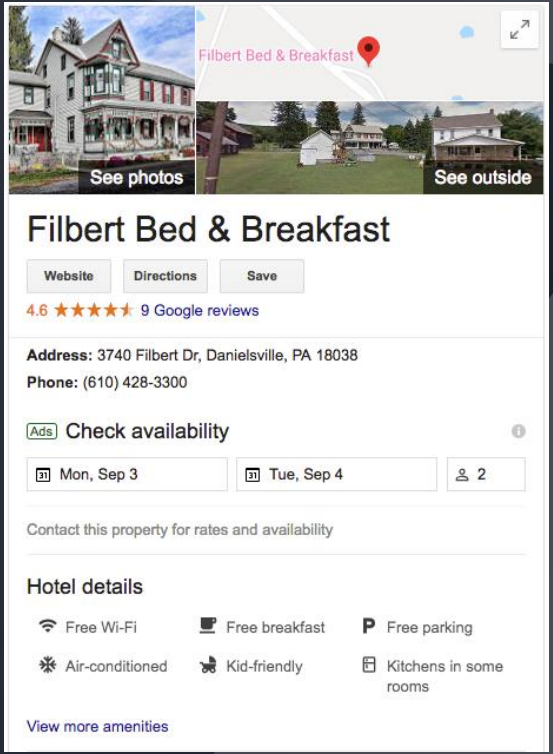
Category Relevance Model – Limitation

Place ID: 840dr621-eaef650f57014ef793231cff5e115f02
Name : Filbert Bed & Breakfast
Category : Transportation Service
Model Prediction : 0.95

All categories for this POI:

Category	Description	Prediction
400-4100-0041	Transportation Service	0.95
700-7400-0284	Wedding Services and Bridal Studio	0.92
500-5000-0000	Hotel or Motel	0.71
500-5100-0058	Bed and Breakfast	0.09

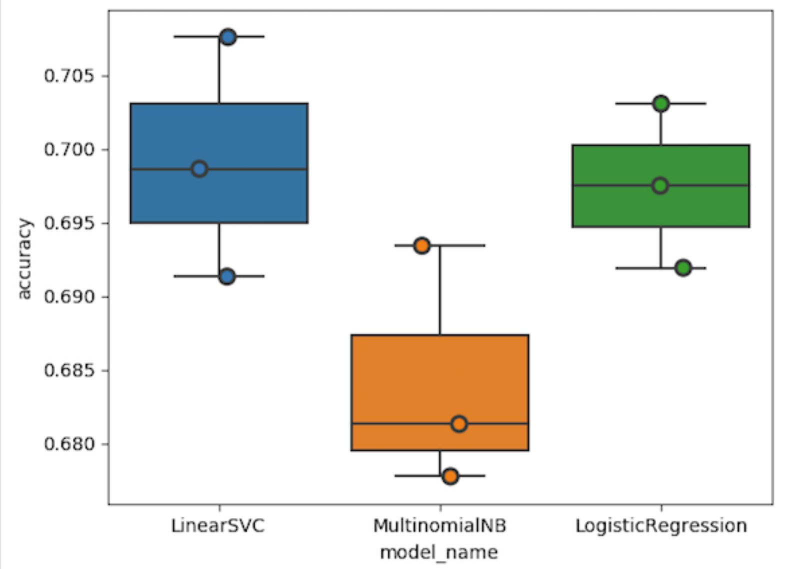
Category from POI Name Prediction:
- Hotel / Lodging = .87



Credit : Matthew Tetkoskie

Category Prediction from Place Name

Q : Can we predict place category based on place (POI) name?
A : Yes, With approximately 70% accuracy.



Category	Precision	Recall
restaurant	64%	62%
coffee shop	84%	59%
nightlife-entertainment	34%	56%
theatre,music,culture	50%	46%
museum	86%	78%
church / reglious place	92%	85%
hotel/motel	71%	67%
lodging	60%	54%
drugstore / pharmacy	98%	92%
hardware/house/garden	78%	78%
bookstore	81%	70%
hair and beauty	92%	80%
car repair	88%	83%
sports facility-venue	58%	72%
Average values	74%	70%

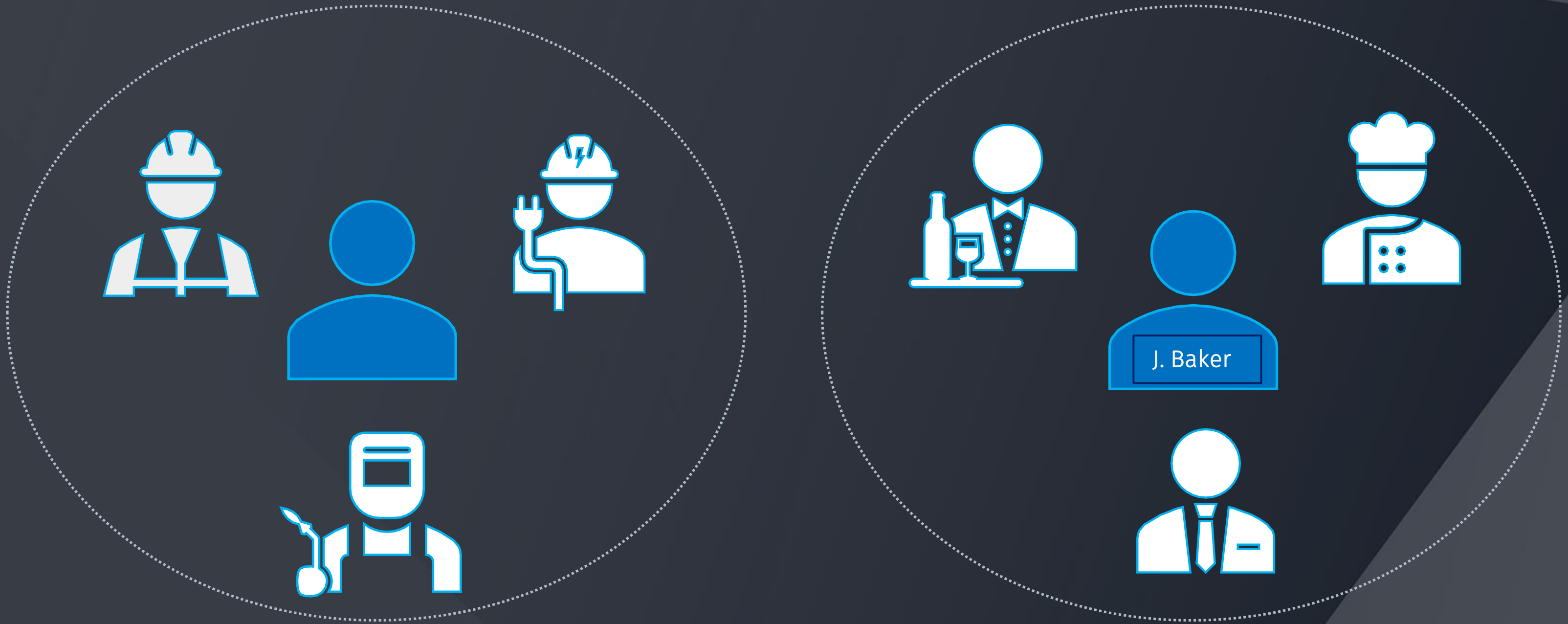
Credit : Matthew Tetkoskie

Analogy : Guess his occupation

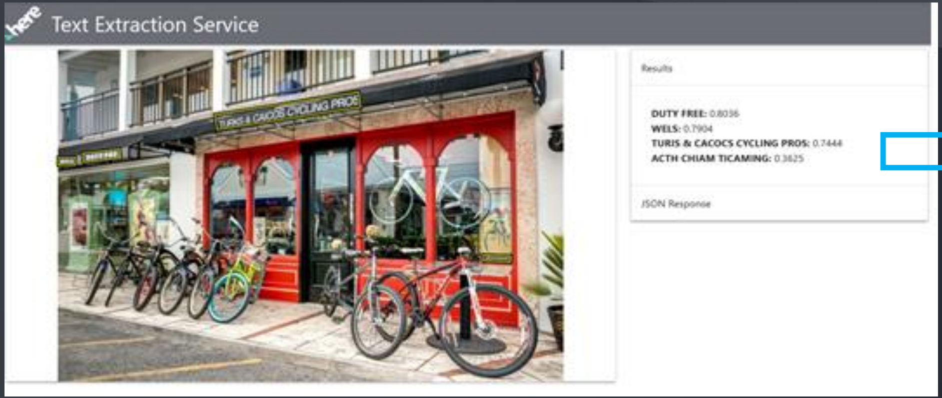
Analogy : Guess his occupation with name tag

Analogy : Uniform helps

Analogy : Without uniform but his neighbors

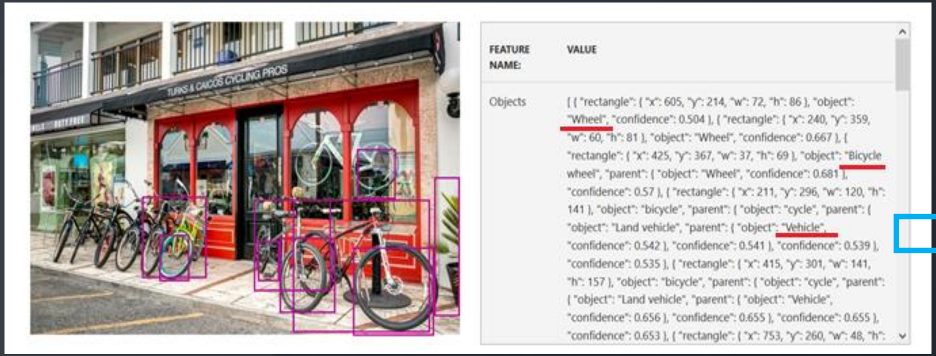


Uniforms ←→ Extracted text & identified objects



Text around storefront

HERE Text Extraction Service



Identified Objects around storefront

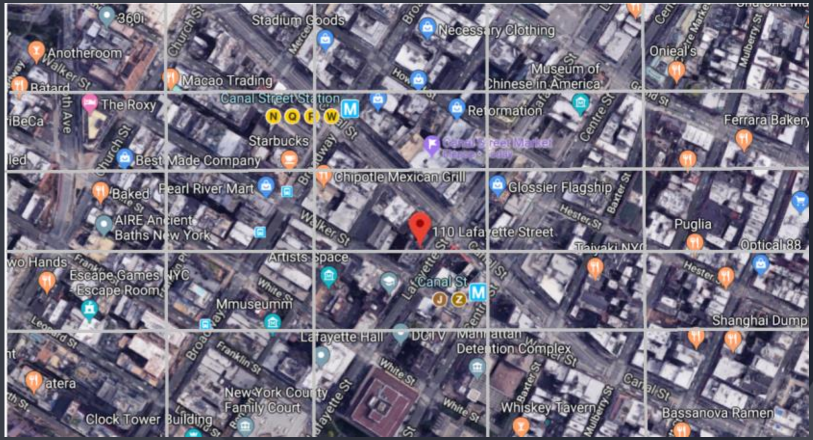
Microsoft Azure Computer Vision API

Thanks to : Xing Fu

Target category prediction using neighboring places' categories

Cat D	Cat D	Cat C	Cat E	Cat F
Cat C	Cat B	Cat A	Cat E	Cat O
Cat C	Cat A	Target	Cat C	Cat A
Cat B	Cat K	Cat B	Cat H	Cat K
Cat A	Cat C	Cat Q	Cat S	Cat A

Cat : category

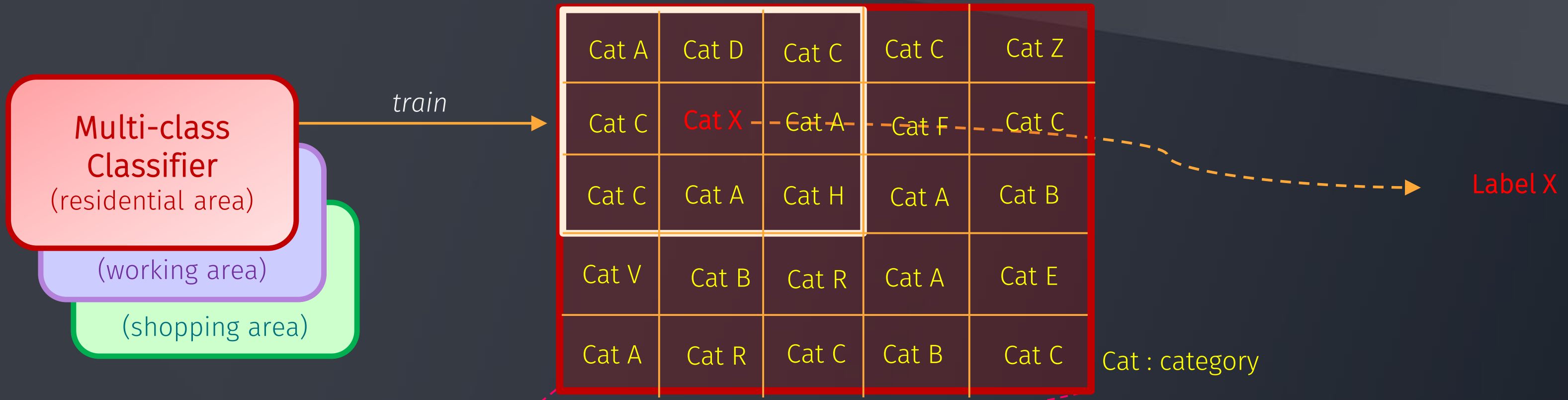


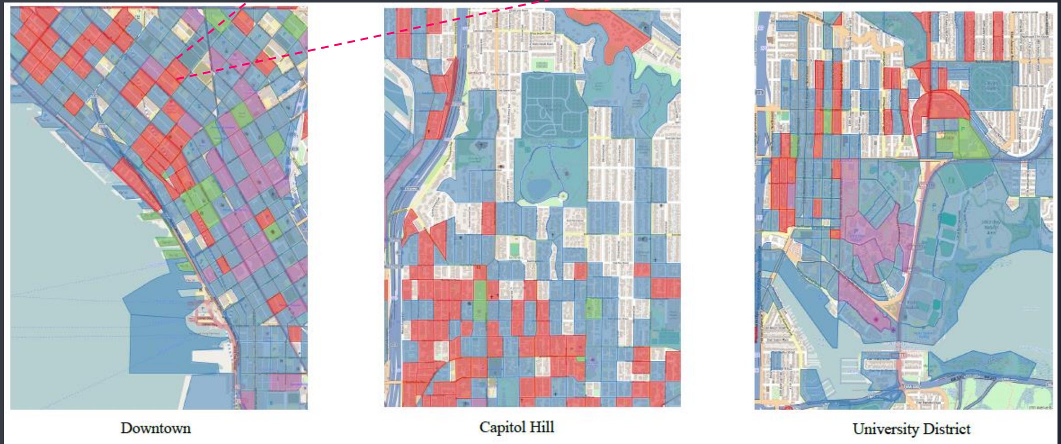
<u>Open questions :</u>

* Representation of neighbors
- matrix, graph, tree

* Prediction Model
- Multiclass Category Classifier
- Estimation on target category by missing entry imputation (e.g. Image inpainting model)

Training data per area type





Red – Residential area
Purple – Working area
Green – Shopping and Going out

image credit : Zack Zhu

Sequential Category Classification for newly discovered place

XXX – YYYY – ZZZZ

{Level 1 – Level 2 – Level 3}

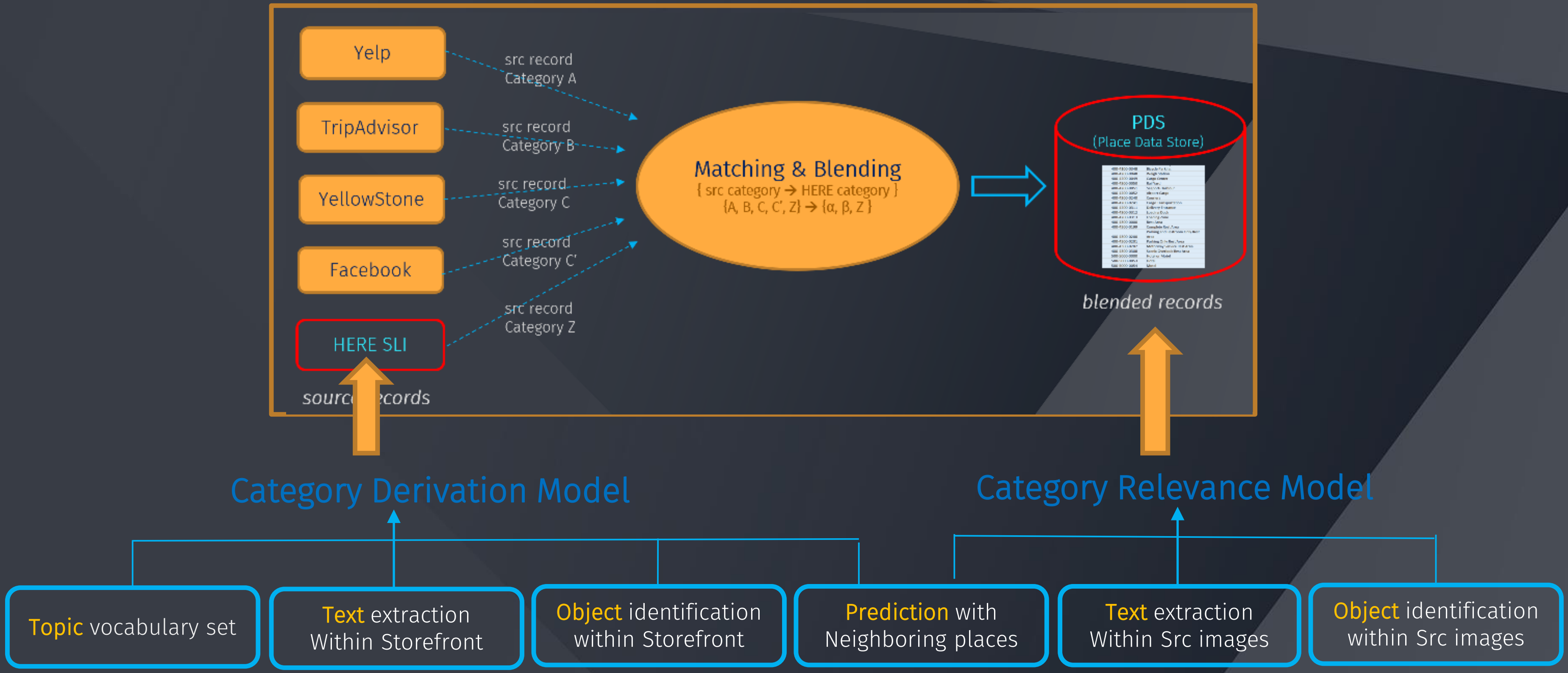
Level 1 Classifier → **Level 2 Classifier** → **Level 3 Classifier**

Level 1 Classifier input:
Topic Vocabulary
(LDA, TF-IDF)
Place Name
Extracted Text
from storefront

Level 2 Classifier input:
Target place's
category prediction
using neighbors

Level 3 Classifier input:
Utilize identified
objects as features

Category model enhancement with additional feature set



Vielen Dank!

Thanks To : Karl Knaub
Matthew Tetkoskie
Charitha H.
Xing Fu
Zack Zhu