

AI from zero to production: lessons from Skyscanner

Konstantin Halachev • SKYSCANNER



Using AI to rank hotels

Every business makes daily expert decisions on **what** to show to **whom** and **when**. Those decisions can often be improved with **AI**

The screenshot shows the Skyscanner website interface for hotel searches. At the top, the Skyscanner logo is on the left, and 'Help', '£ GBP', and 'Account' links are on the right. Below the logo are tabs for 'Flights', 'Hotels', and 'Car Hire'. The search bar contains 'Paris', the dates '19 Apr 2019 - 21 Apr 2019', and '2 guests - 1 room'. A map of Paris is shown on the left with a 'Show map' button. The search results are sorted by 'Best' (1645 results). Filter buttons for 'Price', 'Star rating', 'Guest rating', 'Popular with', 'Meal plan', and 'More filters' are available. The results list shows two hotels:

| Hotel Name | Rating | Reviews | Price | Details |
|--|--------|-----------------------|--------------------|---------------|
| Novotel Suites Paris Montreuil Vincennes | 7.9 | Based on 4509 reviews | 5 prices from £185 | See details → |
| Hôtel L'Interlude | 8.6 | Based on 530 reviews | 7 prices from £163 | See details → |

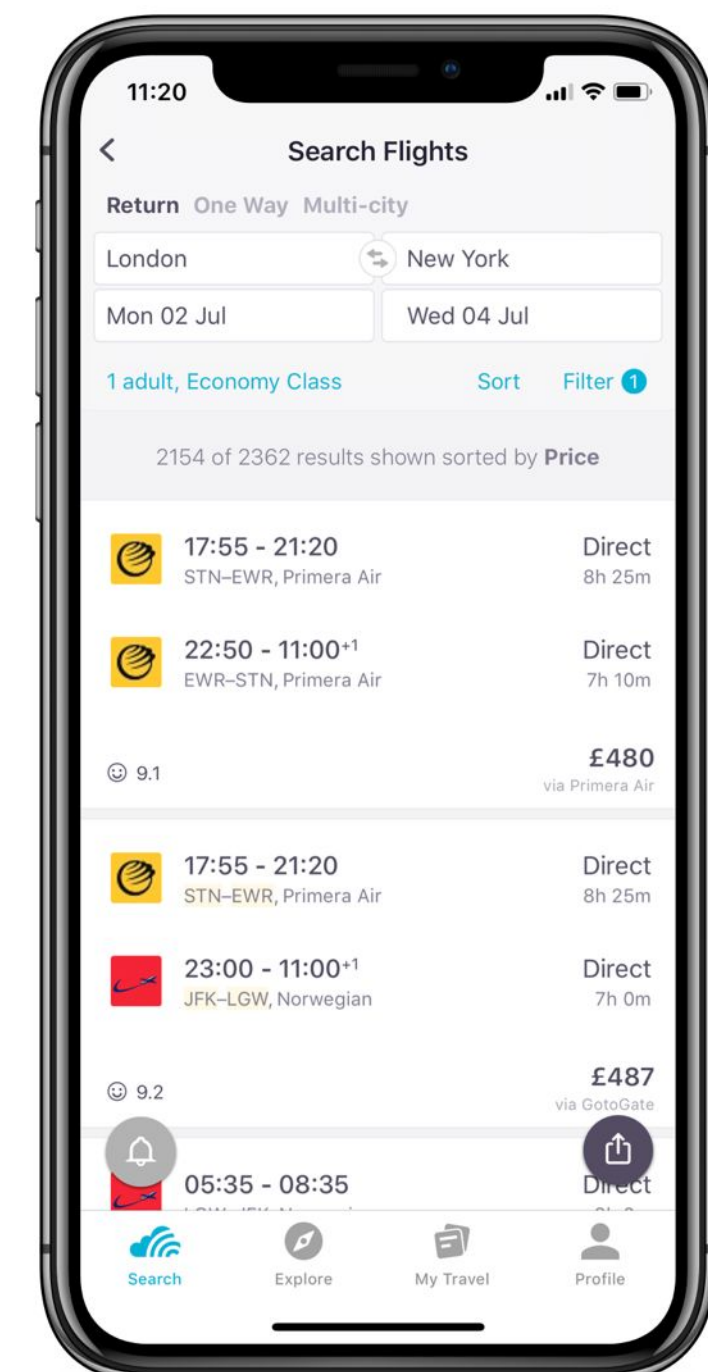
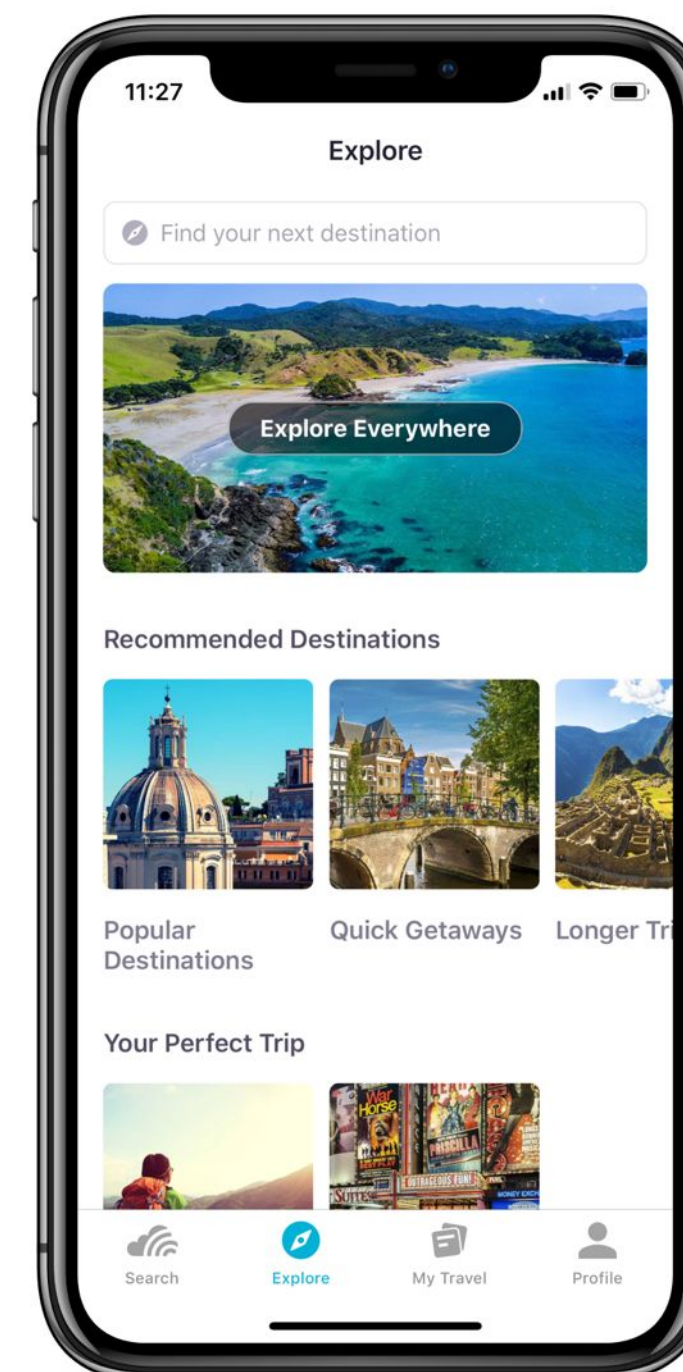
Additional information for each hotel includes amenities (WiFi, Air conditioning, Parking, Pets allowed), location (miles from city centre), and a 'Clean' status.

About Skyscanner



The world's travel search engine

- Powering flights, hotels and car hire.
- Average 70m unique monthly visitors
- Over 70m app downloads



About me

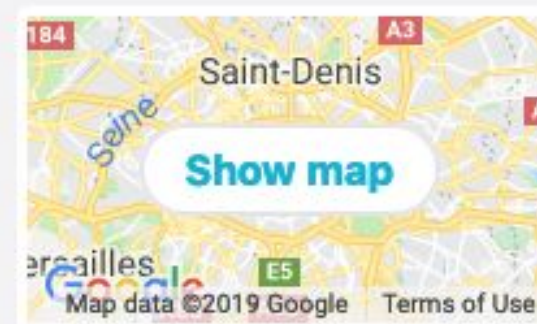
- Konstantin Halachev
- Currently an Engineering manager in Skyscanner
- Previously a Data Scientist
- Previously a PhD in Bioinformatics from MPII

Lessons from ranking hotels

Skyscanner hotel search

[Help](#)[£ GBP](#)[Account](#)[✈ Flights](#)[🏠 Hotels](#)[🚗 Car Hire](#)

Paris 📅 19 Apr 2019 - 21 Apr 2019 🧑 2 guests - 1 room

1645 results sorted by **Best** ⓘSort by **Best** ▼

Filter by

Price ▼

Star rating ▼

Guest rating ▼

Popular with ▼

Meal plan ▼

More filters

Total price for 2 nights, 2 guests, 1 room

All taxes and fees included, except local tax if applicable.



Novotel Suites Paris Montreuil Vincennes

Families

★★★★ 7.9 Based on 4509 reviews

"Clean", "Great room"

WiFi Air conditioning Parking

📍 2.86 miles from city centre

5 prices from

£185

[See details →](#)

Hôtel L'Interlude

★★★★ 8.6 Based on 530 reviews

"Clean", "Fantastic service"

WiFi Air conditioning Pets allowed

📍 2.73 miles from city centre

7 prices from

£163

[See details →](#)

Forest Hill Paris la Villette

Business travellers

Skyscanner's hotel recommendations

1. Business need

Goal of the user: *To find a good hotel*

Goal of Skyscanner: *To offer users good hotels*

Many ways to approach this ...

The screenshot shows the Skyscanner website's hotel search interface. At the top, the Skyscanner logo is on the left, and 'Help', '£ GBP', and 'Account' links are on the right. Below this is a navigation bar with 'Flights', 'Hotels', and 'Car Hire' tabs. The 'Hotels' tab is active. A search bar contains 'Paris', dates '19 Apr 2019 - 21 Apr 2019', and '2 guests - 1 room'. Below the search bar, a map of Paris is shown on the left, and '1645 results sorted by Best' is displayed on the right. A 'Sort by' dropdown menu is set to 'Best'. Below this, a 'Filter by' section includes buttons for 'Price', 'Star rating', 'Guest rating', 'Popular with', 'Meal plan', and 'More filters'. The main content area shows two hotel listings. The first listing is for 'Novotel Suites Paris Montreuil Vincennes', featuring a photo of a hotel room, a 7.9 rating based on 4509 reviews, and a price of £185. The second listing is for 'Hôtel L'Interlude', featuring a photo of the hotel exterior, an 8.6 rating based on 530 reviews, and a price of £163. Both listings include a 'See details' button. At the bottom, a third listing for 'Forest Hill Paris la Villette' is partially visible.

skyscanner

Help £ GBP Account

Flights Hotels Car Hire

Paris 19 Apr 2019 - 21 Apr 2019 2 guests - 1 room

1645 results sorted by Best

Sort by Best

Filter by Price Star rating Guest rating Popular with Meal plan More filters

Total price for 2 nights, 2 guests, 1 room All taxes and fees included, except local tax if applicable.

Novotel Suites Paris Montreuil Vincennes Families

★★★★★ 7.9 Based on 4509 reviews

"Clean", "Great room"

5 prices from £185

See details →

2.86 miles from city centre

Hôtel L'Interlude

★★★★★ 8.6 Based on 530 reviews

"Clean", "Fantastic service"

7 prices from £163

See details →

2.73 miles from city centre

Forest Hill Paris la Villette Business travellers

Skyscanner's hotel recommendations

1. Business need

Goal of the user: *To find a good hotel*

Goal of Skyscanner: *To offer users good hotels*

Many ways to approach this ...

We worked on: *Which hotels are most relevant to a given user, for a given search at a specific time?*

The screenshot shows the Skyscanner website interface for hotel searches. At the top, there's a navigation bar with 'Flights', 'Hotels', and 'Car Hire' tabs. Below this, a search bar contains 'Paris', dates '19 Apr 2019 - 21 Apr 2019', and '2 guests - 1 room'. A map of Paris is visible on the left. The main section displays '1645 results sorted by Best'. Below this, there are filter buttons for 'Price', 'Star rating', 'Guest rating', 'Popular with', 'Meal plan', and 'More filters'. The results list includes two hotels: 'Novotel Suites Paris Montreuil Vincennes' and 'Hôtel L'Interlude'. Each hotel entry shows a photo, name, star rating, guest rating, a quote, amenities, distance from the city center, and a price starting from £185 and £163 respectively. A 'See details' button is present for each hotel.

skyscanner Help £ GBP Account

Flights Hotels Car Hire

Paris 19 Apr 2019 - 21 Apr 2019 2 guests - 1 room

1645 results sorted by Best Sort by Best

Filter by Price Star rating Guest rating Popular with Meal plan More filters

Total price for 2 nights, 2 guests, 1 room All taxes and fees included, except local tax if applicable.

Novotel Suites Paris Montreuil Vincennes Families
★★★★★ 7.9 Based on 4509 reviews
"Clean", "Great room"
WIFI Air conditioning Parking
2.86 miles from city centre
5 prices from **£185**
[See details →](#)

Hôtel L'Interlude
★★★★★ 8.6 Based on 530 reviews
"Clean", "Fantastic service"
WIFI Air conditioning Pets allowed
2.73 miles from city centre
7 prices from **£163**
[See details →](#)

Forest Hill Paris la Villette Business travellers

Which hotels are most relevant to a given user, for a given search at a specific time?

2. Data

THE DATA SCIENCE **HIERARCHY OF NEEDS**

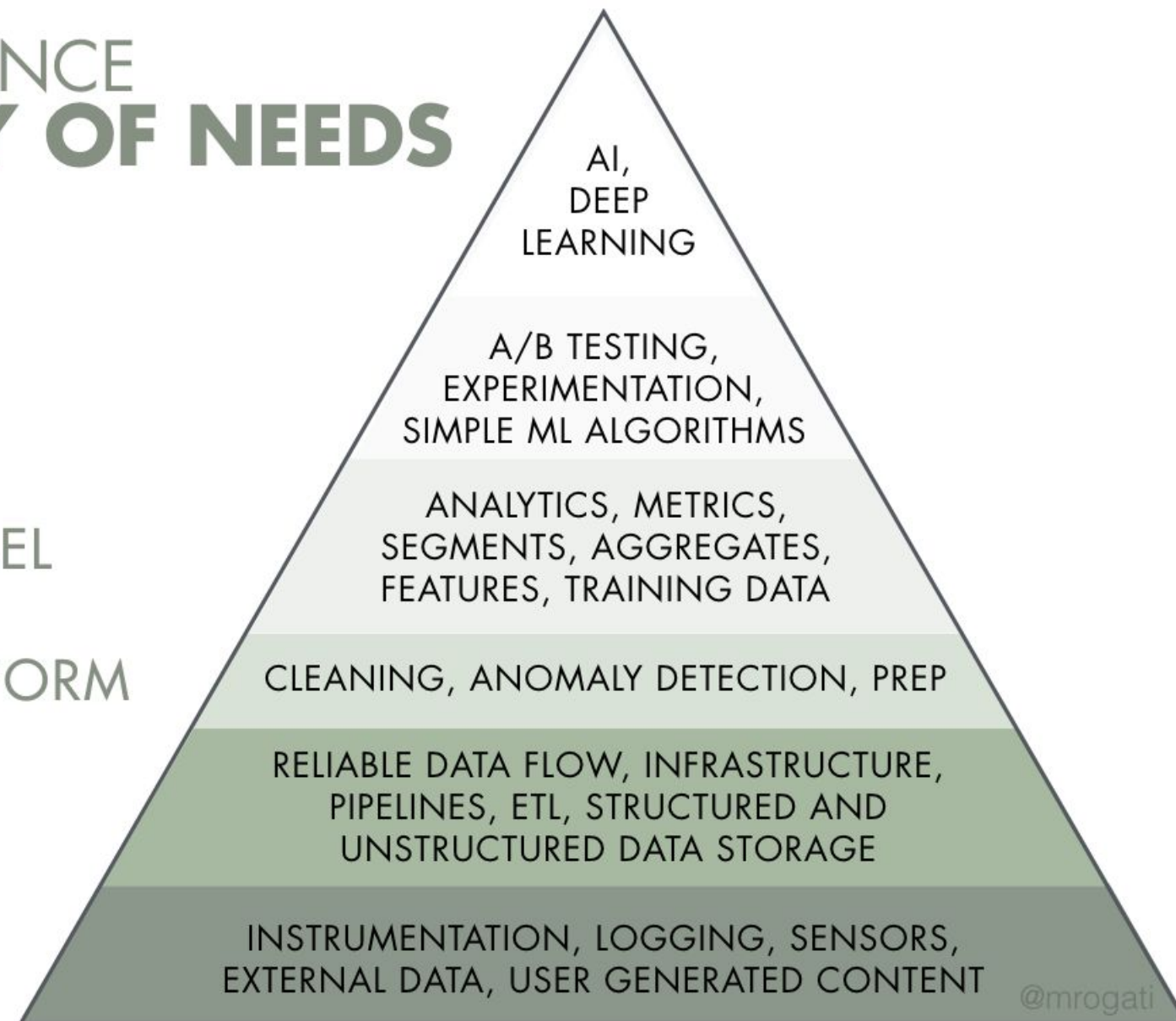
LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT



Which hotels are most relevant to a given user, for a given search at a specific time?

2. Data

THE DATA SCIENCE HIERARCHY OF NEEDS

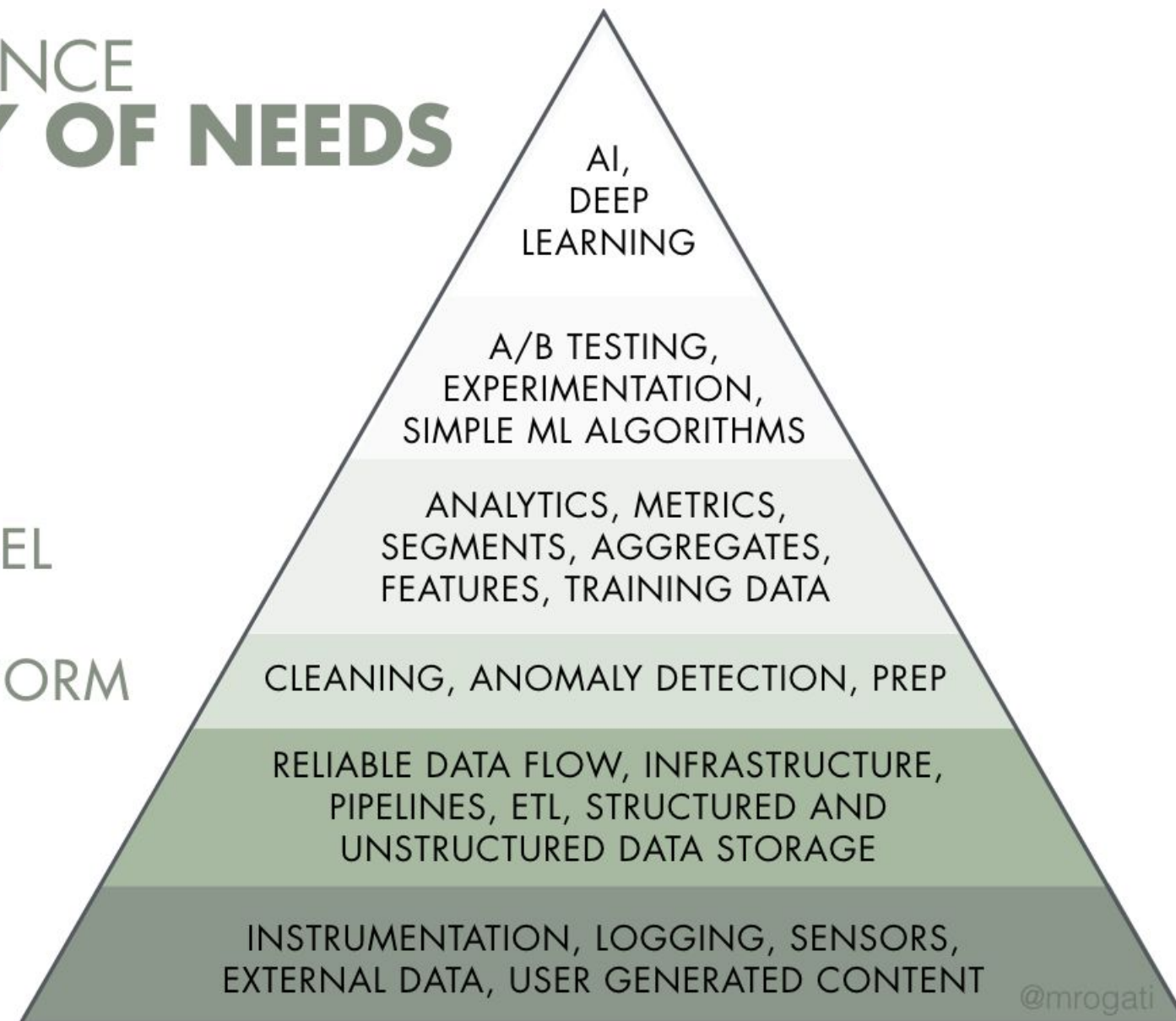
LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT



- Do we have data to answer the question?
 - What have users searched and seen?
 - Do we store the hotels shown?
 - Do we store their positions?
 - Do we store prices, ratings, location, etc?
- Is the data correct?
 - Do we store all searches?
 - Are the searches in the same format?
 - Are there values that do not make sense?
 - Are there outliers?
 - Has the UI or UX changed?
- Do we have enough data?
 - 100 data points are not enough
 - Data for 1 day is not enough
 - Millions of data points may be

THE DATA SCIENCE HIERARCHY OF NEEDS

LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

AI,
DEEP
LEARNING

A/B TESTING,
EXPERIMENTATION,
SIMPLE ML ALGORITHMS

ANALYTICS, METRICS,
SEGMENTS, AGGREGATES,
FEATURES, TRAINING DATA

CLEANING, ANOMALY DETECTION, PREP

RELIABLE DATA FLOW, INFRASTRUCTURE,
PIPELINES, ETL, STRUCTURED AND
UNSTRUCTURED DATA STORAGE

INSTRUMENTATION, LOGGING, SENSORS,
EXTERNAL DATA, USER GENERATED CONTENT

@mrogati

THE DATA SCIENCE HIERARCHY OF NEEDS

The diagram is a pyramid divided into six horizontal layers, each representing a level of the data science hierarchy. The layers are labeled on the left and contain specific tasks or concepts within the pyramid.

| Level | Tasks/Concepts |
|------------|--|
| 1 (Bottom) | INSTRUMENTATION, LOGGING, SENSORS, EXTERNAL DATA, USER GENERATED CONTENT |
| 2 | RELIABLE DATA FLOW, INFRASTRUCTURE, PIPELINES, ETL, STRUCTURED AND UNSTRUCTURED DATA STORAGE |
| 3 | CLEANING, ANOMALY DETECTION, PREP |
| 4 | ANALYTICS, METRICS, SEGMENTS, AGGREGATES, FEATURES, TRAINING DATA |
| 5 | A/B TESTING, EXPERIMENTATION, SIMPLE ML ALGORITHMS |
| 6 (Top) | AI, DEEP LEARNING |

LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

@mrogati

- # 80%

Assume we have a new ranking algorithm

3. Evaluate

THE DATA SCIENCE HIERARCHY OF NEEDS

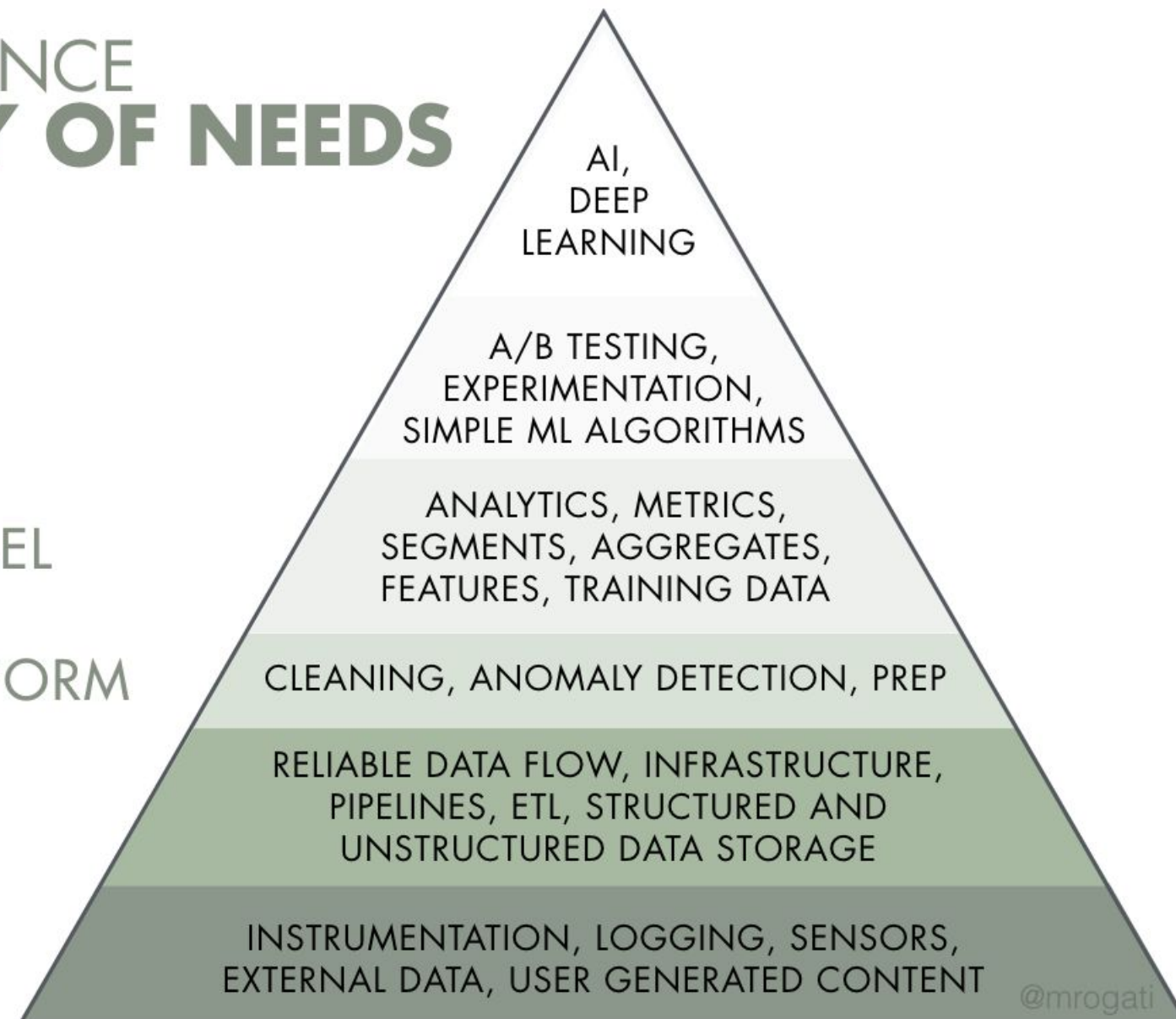
LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT



- How do we put this algorithm in front of users?
- How do we compare against the previous solution?
- What do we need to keep track of?
- Did we improve significantly?

4. Ranking algorithm

Learning to rank hotels

skyscanner [Help](#) [£ GBP](#) [Account](#)


[Flights](#) [Hotels](#) [Car Hire](#)

Paris 19 Apr 2019 - 21 Apr 2019 2 guests - 1 room

1645 results sorted by **Best** [i](#) Sort by **Best** ▼

Filter by [Price ▼](#) [Star rating ▼](#) [Guest rating ▼](#) [Popular with ▼](#) [Meal plan ▼](#) [More filters](#)

Total price for 2 nights, 2 guests, 1 room All taxes and fees included, except local tax if applicable.



Novotel Suites Paris Montreuil Vincennes [Families](#)

★★★★★ **7.9** Based on 4509 reviews


"Clean", "Great room"

WiFi Air conditioning Parking

2.86 miles from city centre

5 prices from **£185**

[See details →](#)



Hôtel L'Interlude

★★★★ **8.6** Based on 530 reviews


"Clean", "Fantastic service"

WiFi Air conditioning Pets allowed

2.73 miles from city centre

7 prices from **£163**

[See details →](#)



Forest Hill Paris la Villette [Business travellers](#)

A few properties that domain experts use to order hotels

- Room price per night
- Hotel location
- User Rating
- Hotel Stars
- Hotel Amenities
- City
- Pictures


4. Ranking algorithm

4.1 Problem representation


1934 results sorted by **Best** ⓘ

Filter by **Price** ▼ More filters

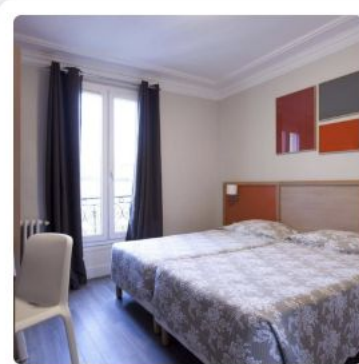
Total price for 2 nights, 2 guests, 1 room
All taxes and fees included, except local tax if applicable.




Hôtel L'Interlude
★★★★ 8.6
2.73 miles from city centre
6 prices from **£120**



Novotel Suites Paris Montreuil Vincennes
★★★★★ 7.9
Families
2.86 miles from city centre
5 prices from **£152**



Est Hotel
★★★★ 7.8
Couples
1.17 miles from city centre
5 prices from **£141**



Forest Hill Paris la Villette
★★★★★ 7.9

What the user saw and did not choose

What the user saw and did not choose

What the user saw and chose

What the user did not see

4. Ranking algorithm

4.2 Exploration space

Simple vs Complex models

- Amount of data
- Need for interpretability
- Domain knowledge (features)



Gradient boosted trees

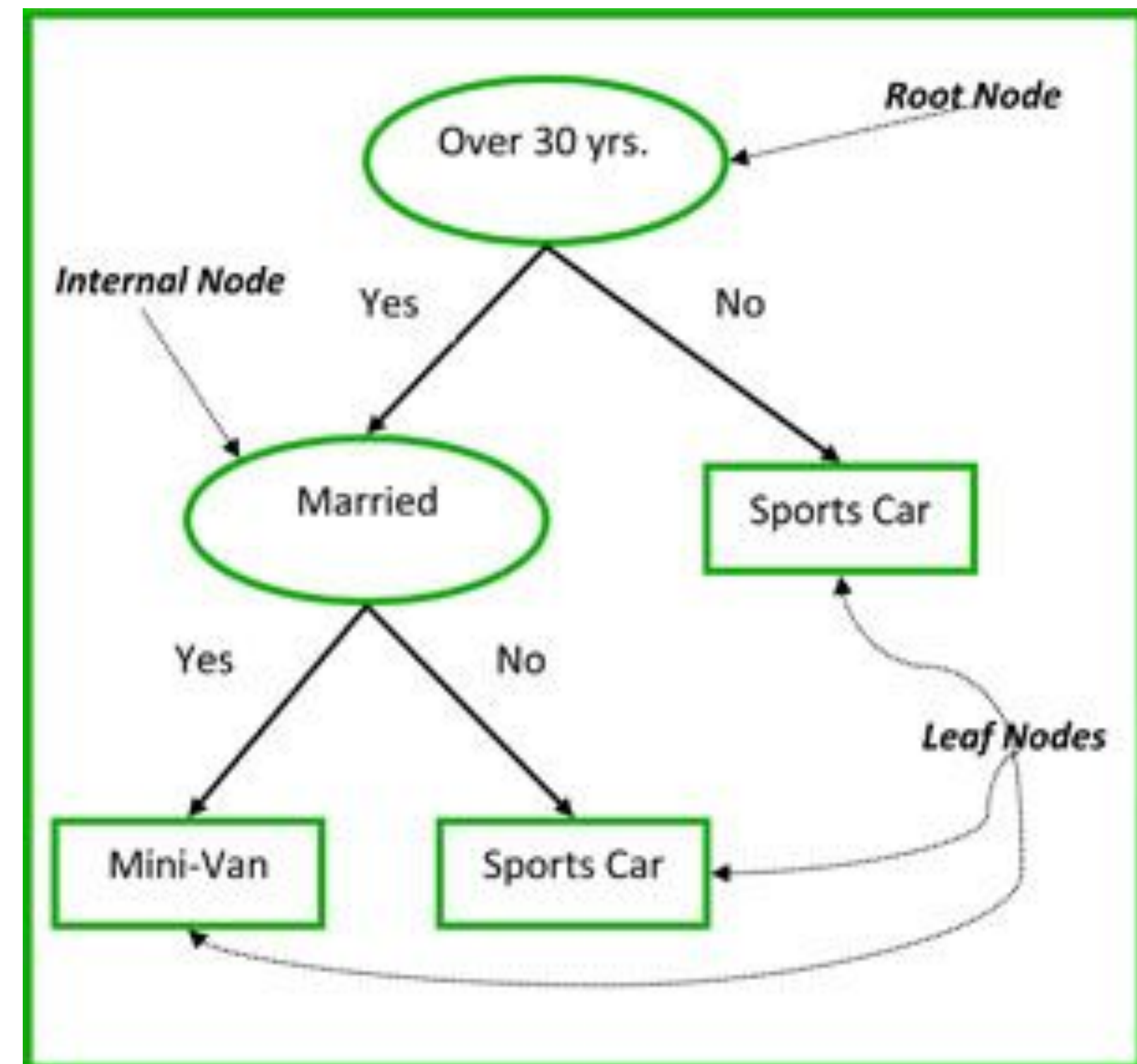
4. Ranking algorithm

4.2 Exploration space



DECISION TREE

Decision trees



4. Ranking algorithm

4.3 Evaluation



Is there a ground truth?



Can we reproduce the experiment?



Can we test with real users?



Can we test offline?

With assumption for resorting stability

Comparison of multiple models

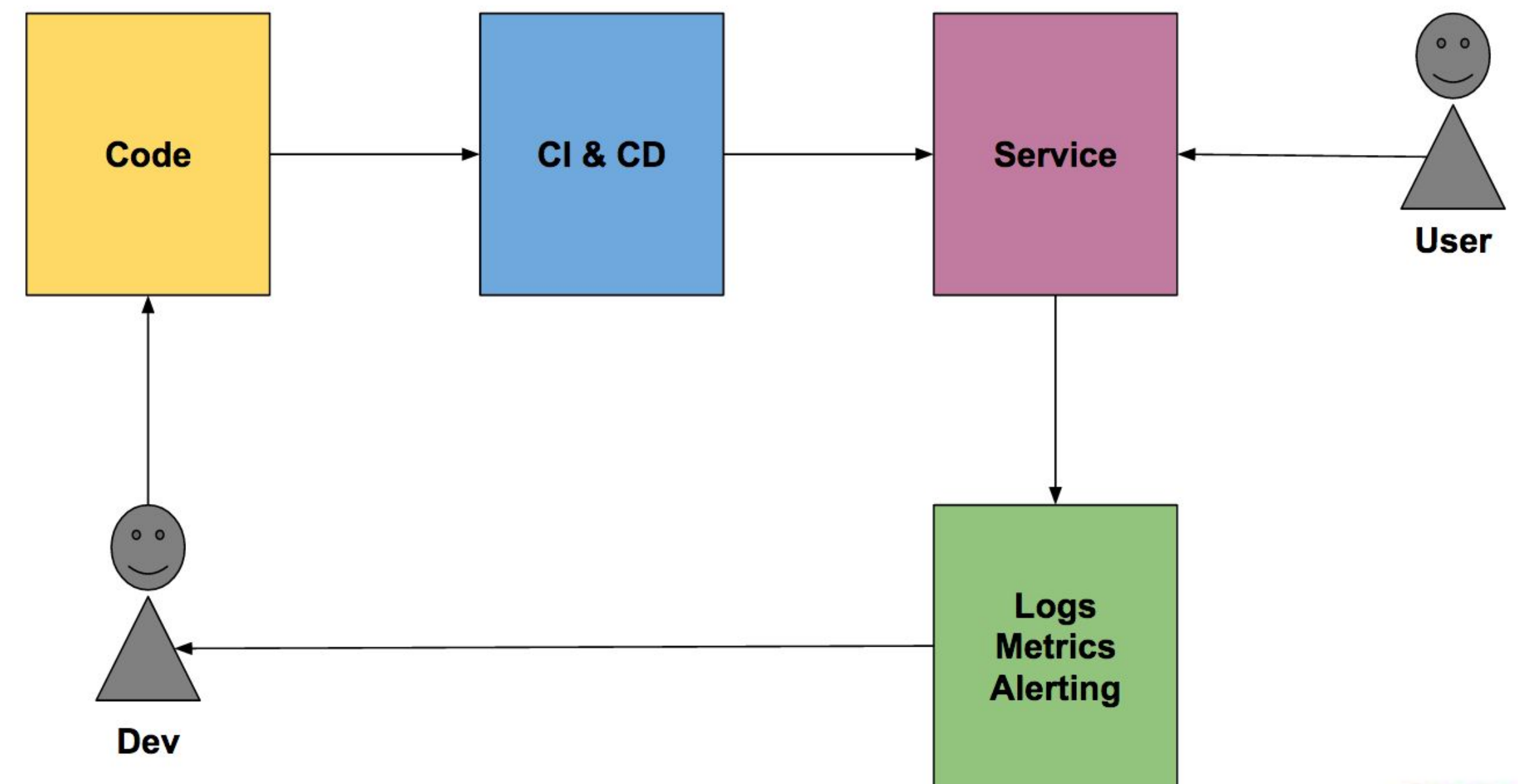
Great!

We have a model, now what?

SUBTITLE

Getting to production

- Continuous integration (CI)
- Continuous Deployment (CD)
- Testing
- Monitoring



SUBTITLE

After being in
prod. It's not
over

- Effect of latency
- Bias (e.g. self-fulfilling prophecies)
- Updating models

Hidden Technical Debt in Machine Learning Systems

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips
{dsculley, gholt, dgg, edavydov, toddphillips}@google.com
Google, Inc.

Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-François Crespo, Dan Dennison
{ebner, vchaudhary, mwyoung, jfcrespo, dennison}@google.com
Google, Inc.

Abstract

Machine learning offers a fantastically powerful toolkit for building useful complex prediction systems quickly. This paper argues it is dangerous to think of these quick wins as coming for free. Using the software engineering framework of *technical debt*, we find it is common to incur massive ongoing maintenance costs in real-world ML systems. We explore several ML-specific risk factors to account for in system design. These include boundary erosion, entanglement, hidden feedback loops, undeclared consumers, data dependencies, configuration issues, changes in the external world, and a variety of system-level anti-patterns.

Results

Start simple and iterate

Lean 3-step approach

1. A/A' : **Ensure no harm**

Start simple and iterate

Lean 3-step approach

1. A/A' : **Ensure no harm**
2. A/B with simple model: **Success**



Start simple and iterate

Lean 3-step approach

1. A/A` : **Ensure no harm**
2. A/B with simple model: **Success**
3. A/B with more complex models: **More success**

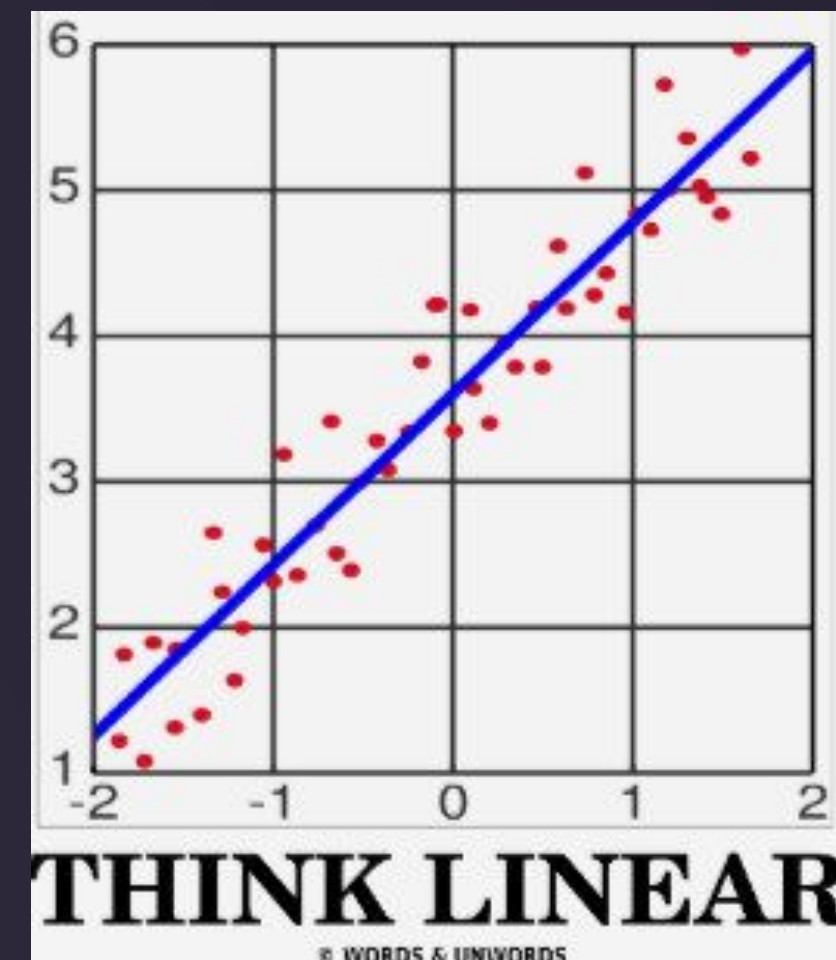
One last thing

A few properties that domain experts use to order hotels

- Room price per night
- Hotel location
- User Rating
- Hotel Stars
- Hotel Amenities
- City
- Pictures

These properties were enough to offer users better hotels than a domain expert

- Price
- ~~Location~~
- Rating
- Stars
- ~~Amenities~~
- ~~City~~
- ~~Picture~~



To summarize

- AI holds a lot of promise for businesses
- Expertise, tools and materials are available
- Even simple solutions can beat domain experts. Start simple.

Thank you.

Konstantin Halachev • SKYSCANNER

