Best Time 4 Contact

Data Science @ NOS

18

Data Scientists

250+

GB of data/day

4.1+

TB of RAM

Established

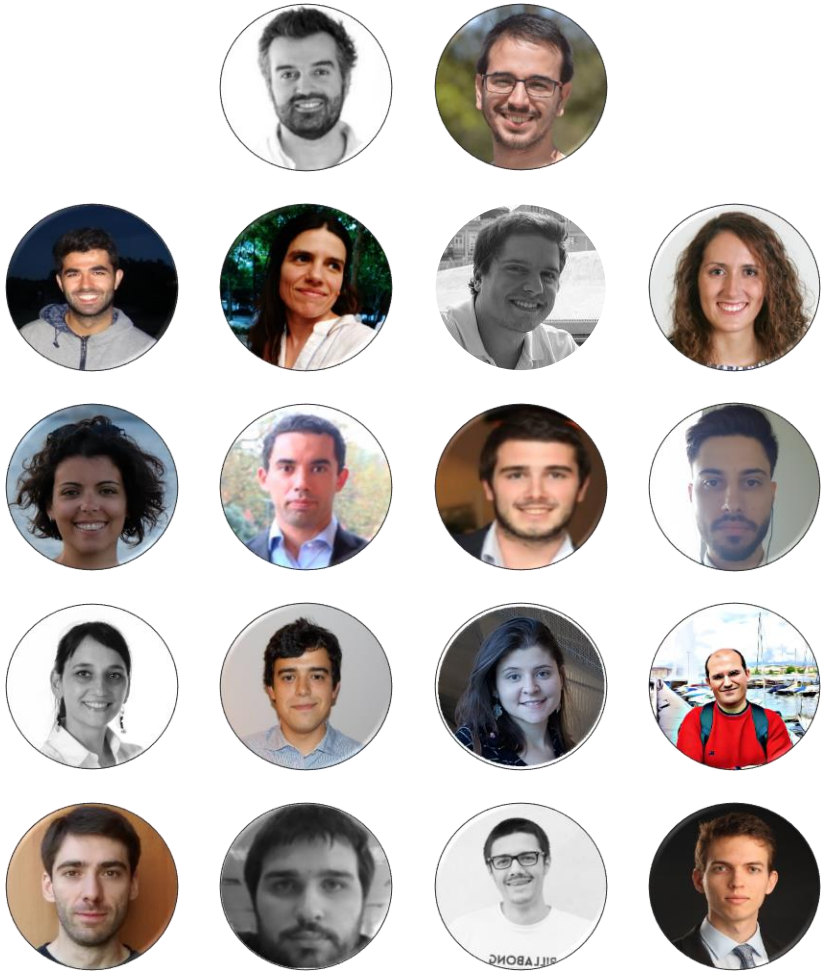
- Supervised Learning
- Oper. Research & Simulation
- Recommender Systems
- Unsupervised Learning

PoC

- Network / Graph Analysis
- NLP / Speech
- Data Vizualization / GIS

Roadmap

- Reinforcement Learning
- Computer Vision



Best Time 4 Contact

NOS needs to call customers to communicate discounts, contractual terms, resolve technical issues, etc.

Calling at the wrong time is inconvenient for the client and affects the likelihood of success of further calls.

Current dialling strategy is left to partners at call centers and is often rules/quarentines based.

>12M

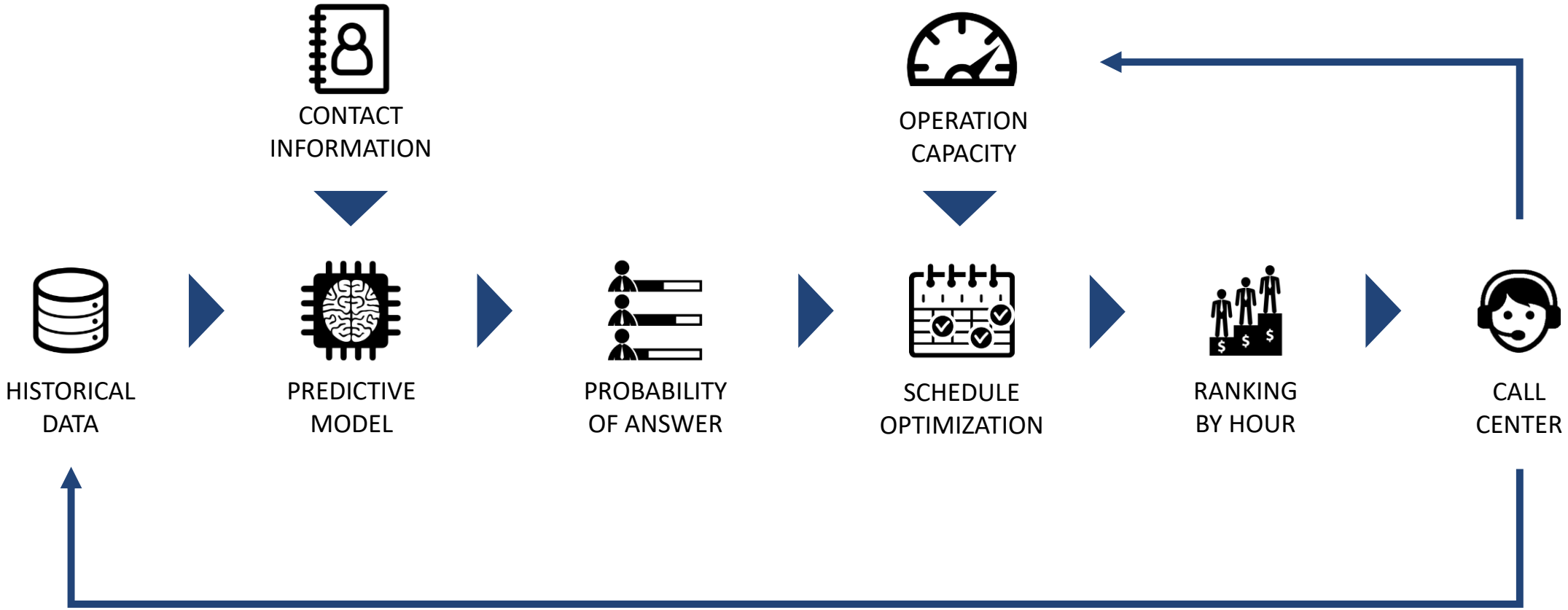
Calls since start of 2019

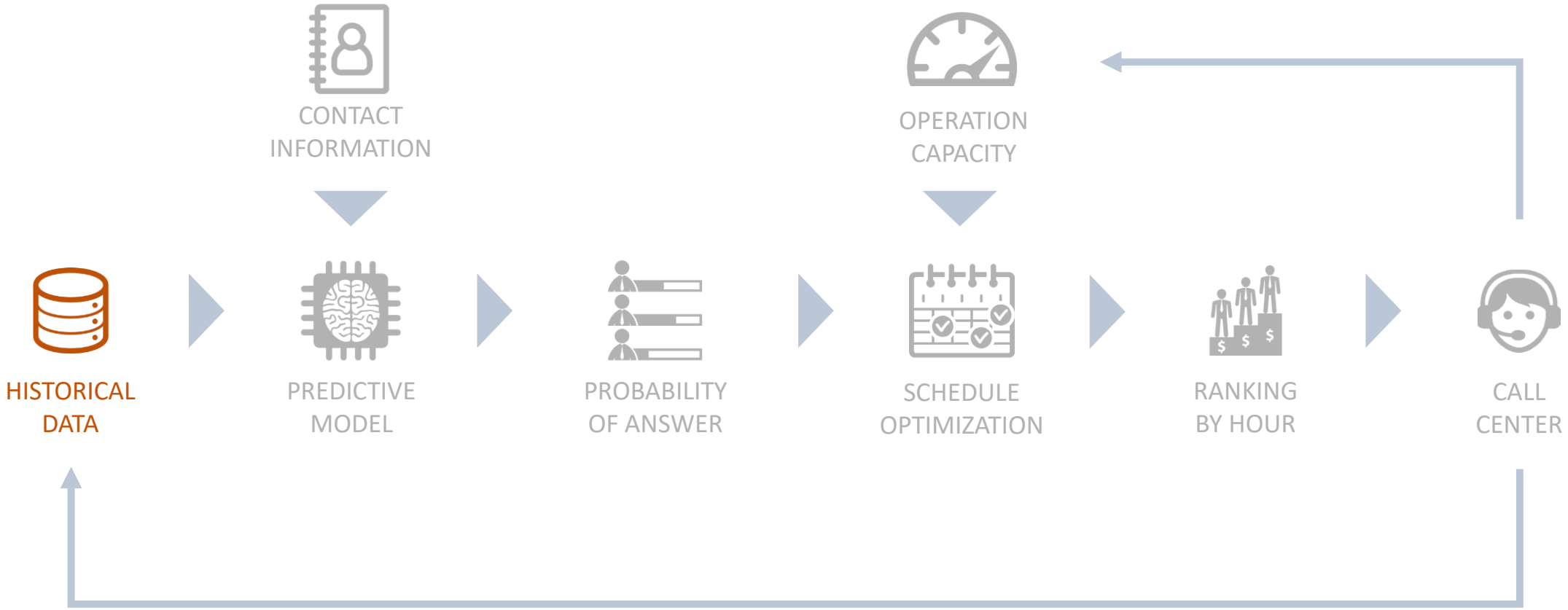
15.9% (1)

Calls answered

Use ML to learn from existing data and inform each partner when to contact each client.

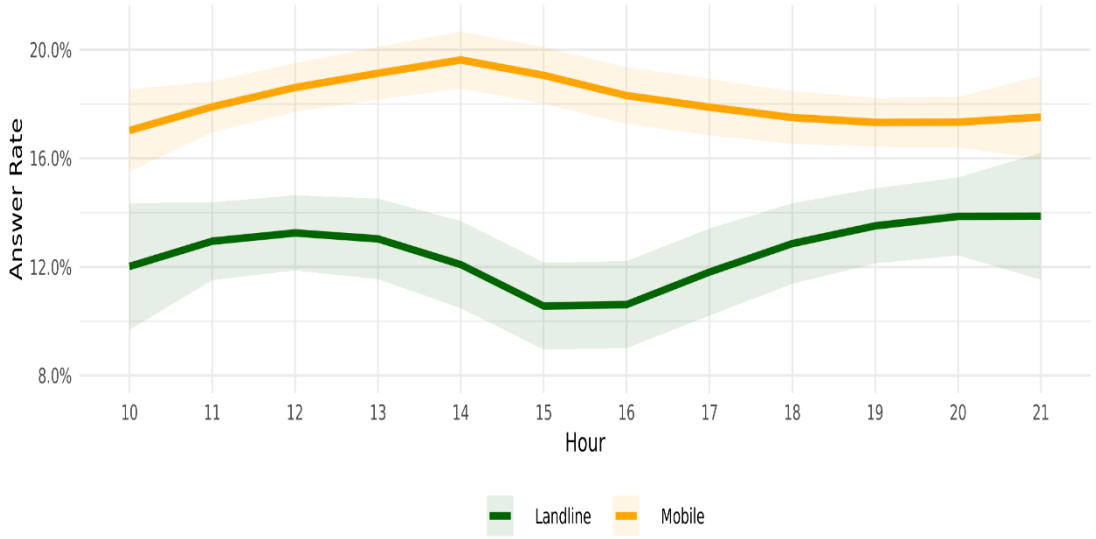
(1) On one of the largests call center operations.


CONTACT INFORMATION
OPERATION CAPACITY
HISTORICAL DATA
PREDICTIVE MODEL
PROBABILITY OF ANSWER
SCHEDULE OPTIMIZATION
RANKING BY HOUR
CALL CENTER


CONTACT INFORMATION
OPERATION CAPACITY
HISTORICAL DATA
PREDICTIVE MODEL
PROBABILITY OF ANSWER
SCHEDULE OPTIMIZATION
RANKING BY HOUR
CALL CENTER

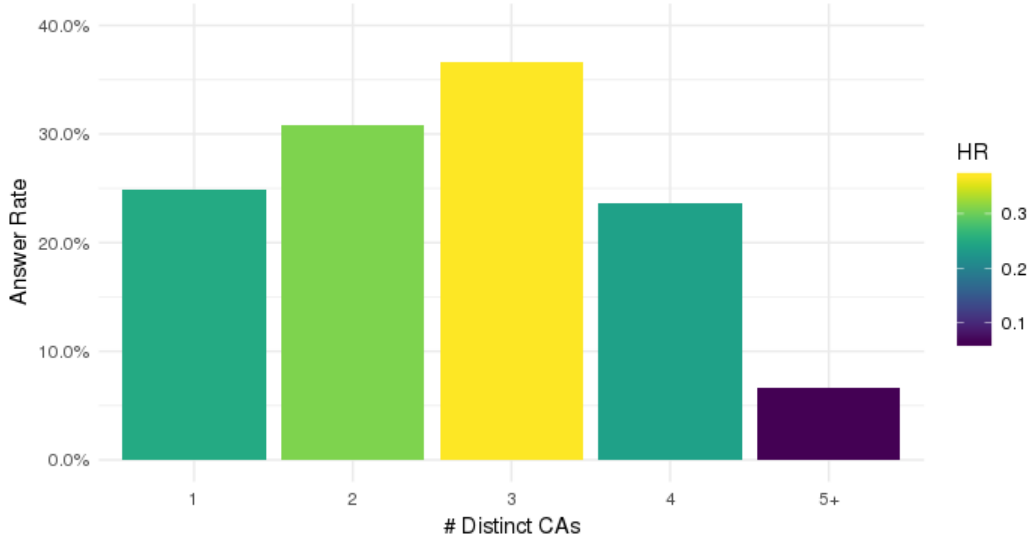
Data Analysis

Answer rate by MSISDN type over time



- Calling mobile phones yields better results on average.
- The best time to contact a mobile phone is is not the same as a landline phone.

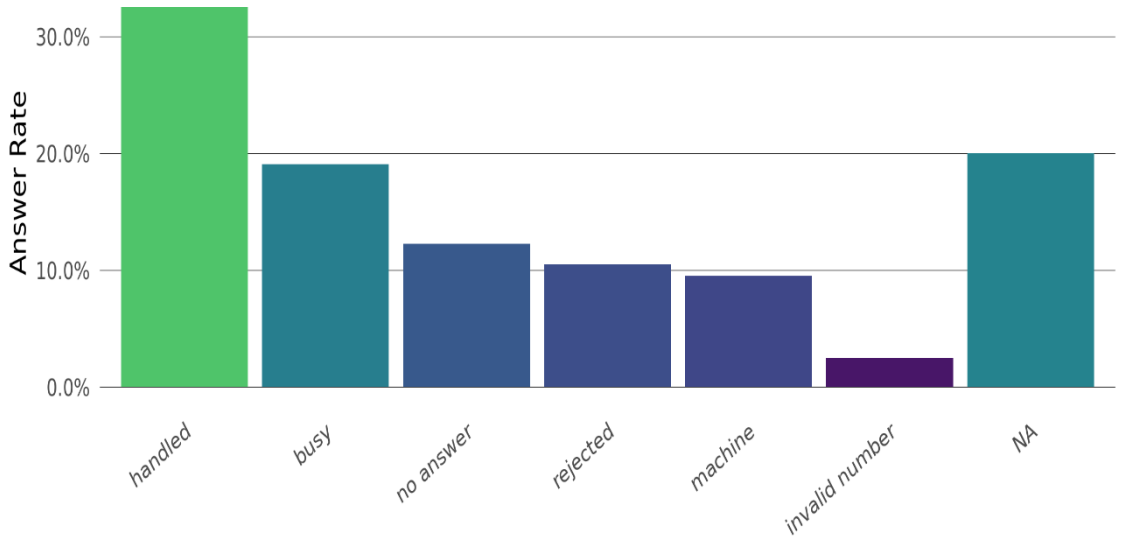
Answer rate by number of shared CAs



- Some MSISDNs are associated with multiple clients.
- Some are default values, fake numbers, etc.

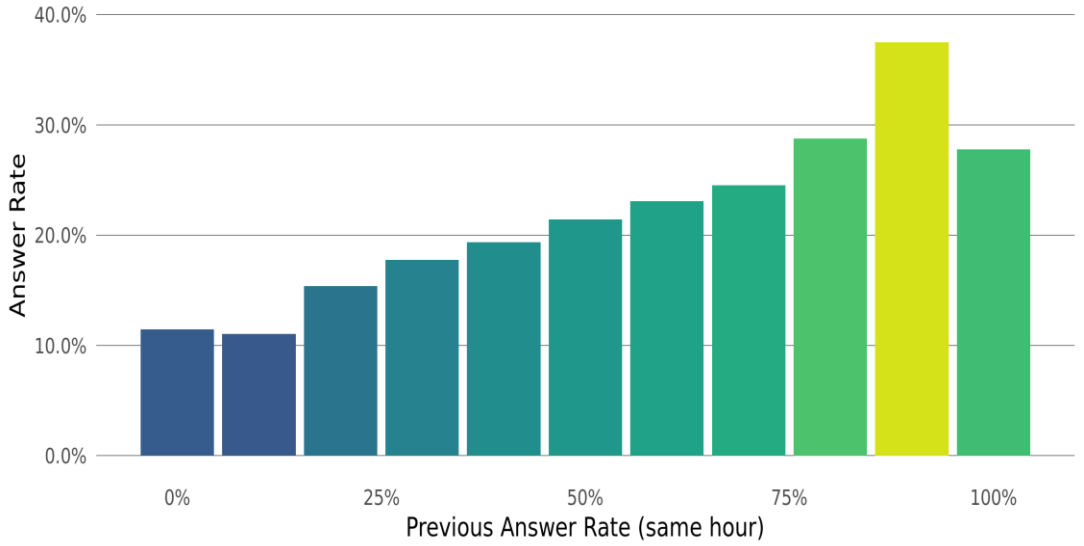
Data Analysis

Answer Rate by Previous Contact result



- Previous call results are important to predict future calls.
- If the client already answered once the likelihood of answering again goes up to over 30%.

Answer Rate by previous rate (same hour)



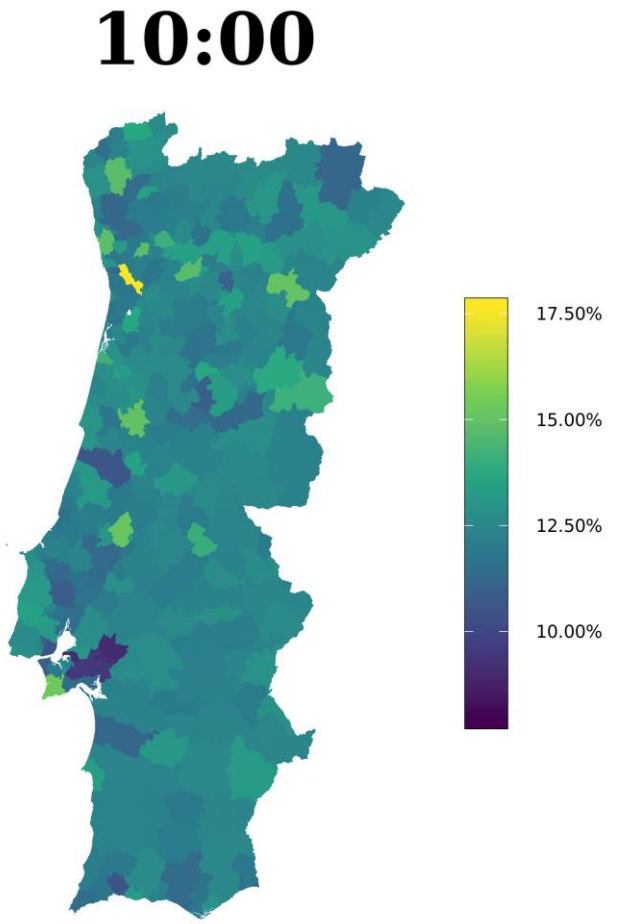
- If the client answers the phone frequently at a specific time of day, that is a good time to try and call again.

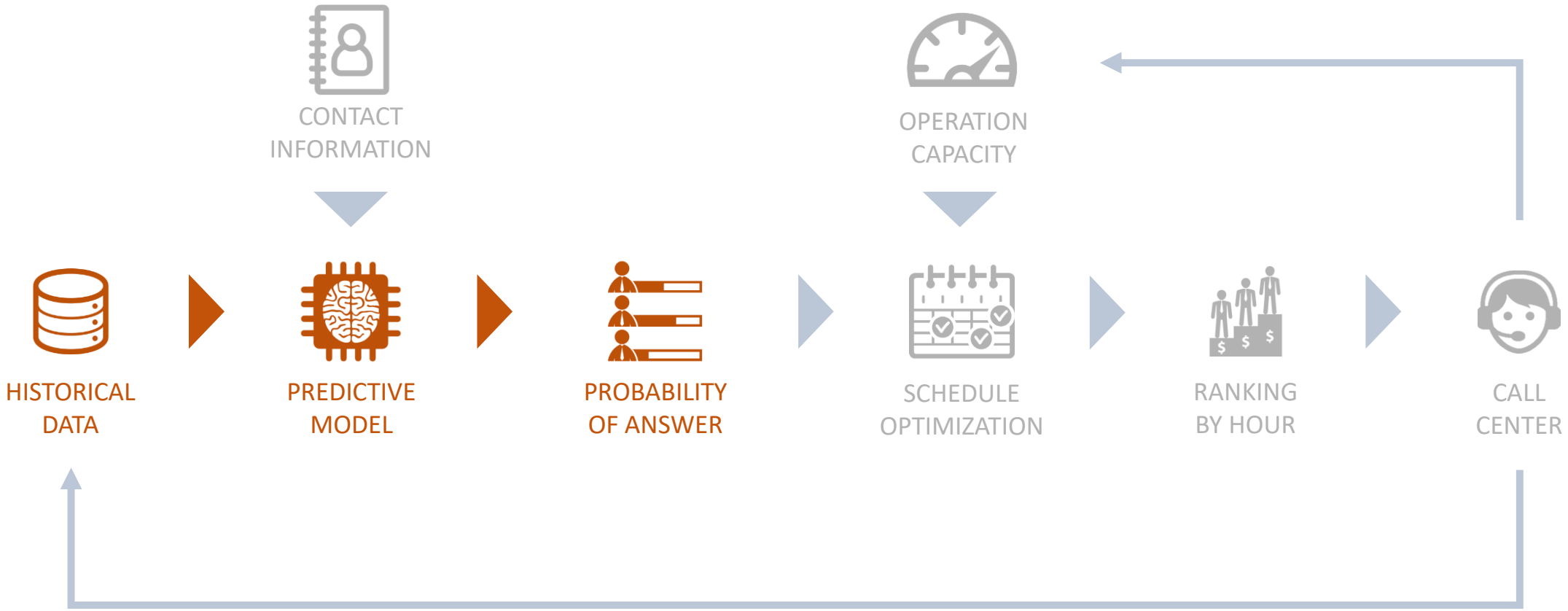
Data Analysis

There is more availability between 12:00 – 13:00 and then again after 18:00.

Each region has a specific behaviour over time.

We were hoping to observe larger differences between costal and inland zones.




CONTACT INFORMATION
OPERATION CAPACITY
HISTORICAL DATA
PREDICTIVE MODEL
PROBABILITY OF ANSWER
SCHEDULE OPTIMIZATION
RANKING BY HOUR
CALL CENTER

Predictive Model

Train

x_hour	x_answer_perc	x_mobile_type	x_prev_result	x_num_calls_hour	…	y_target
10	30%	T	Handled	3	…	T
12	0%	F	Busy	1	…	F
22	75%	T	Handled	2	…	F

Apply

x_hour	x_answer_perc	x_mobile_type	x_prev_result	x_num_calls_hour	…	y_target
10	30%	T	Handled	5	…	??
11	30%	T	Handled	2	…	??
12	30%	T	Handled	1	…	??
…	…	…	…	…	…	…
21	30%	T	Handled	1	…	??

- Manipulating feature values allows us to simulate what-if scenarios and make predictions for future calls.
- Must be careful not to invalidate other feature values (e.g. data-leakage)

Predictive Model

Call History
- Time and Day of contact
- Outcome
- Client/MSISDN information
- Reschedules

85

Contractual Information
- Account Type and Age
- Services and Technology
- Preferential Contact Hours
- Revenue
- ….

26

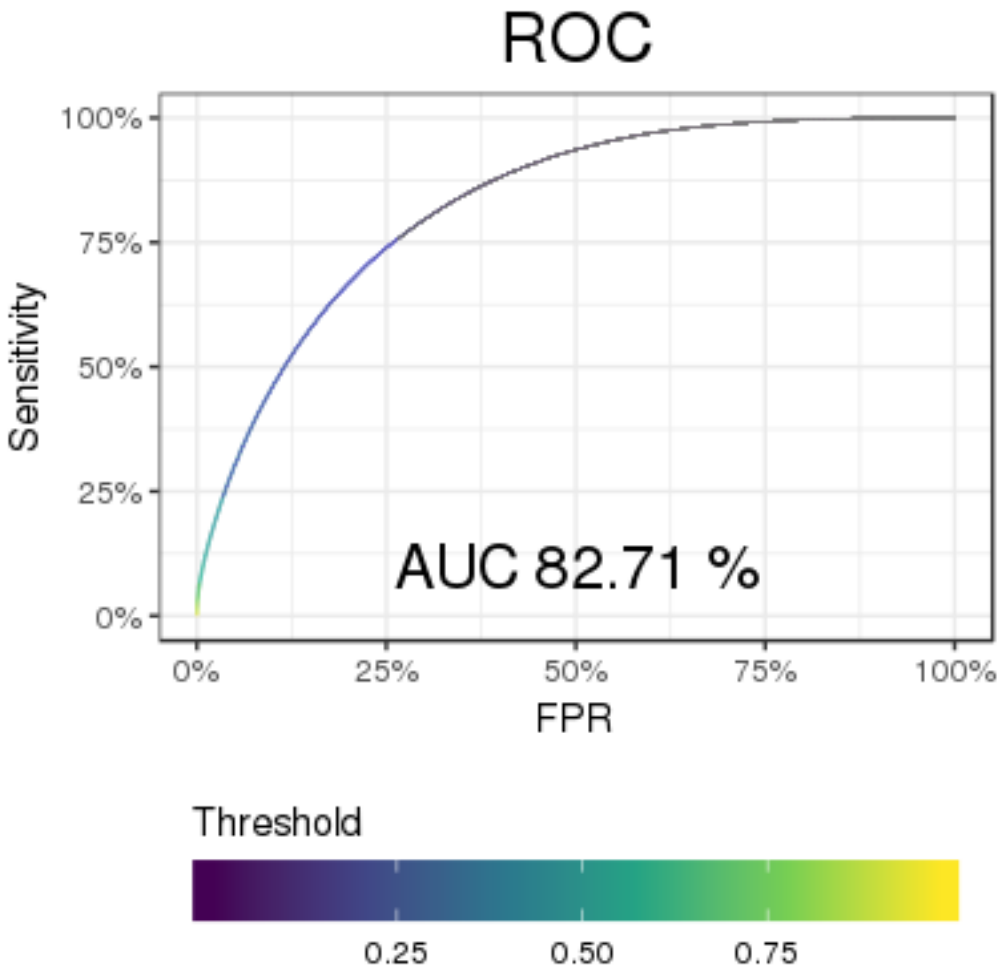
111
Features

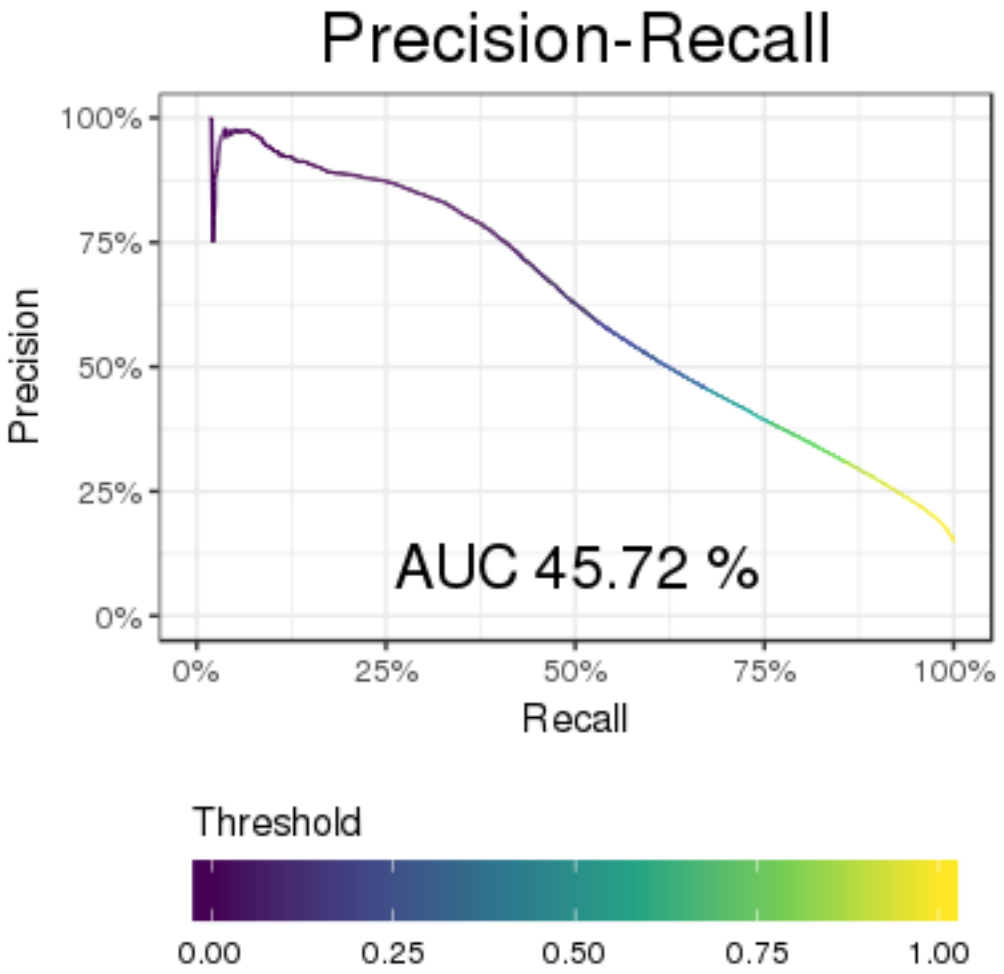
Call Profiles
- By client, MSISDN or both
- Overall vs last 30 days
- # calls, % of answers, avg. call duration, …

Target Encoding
- Specific for each hour

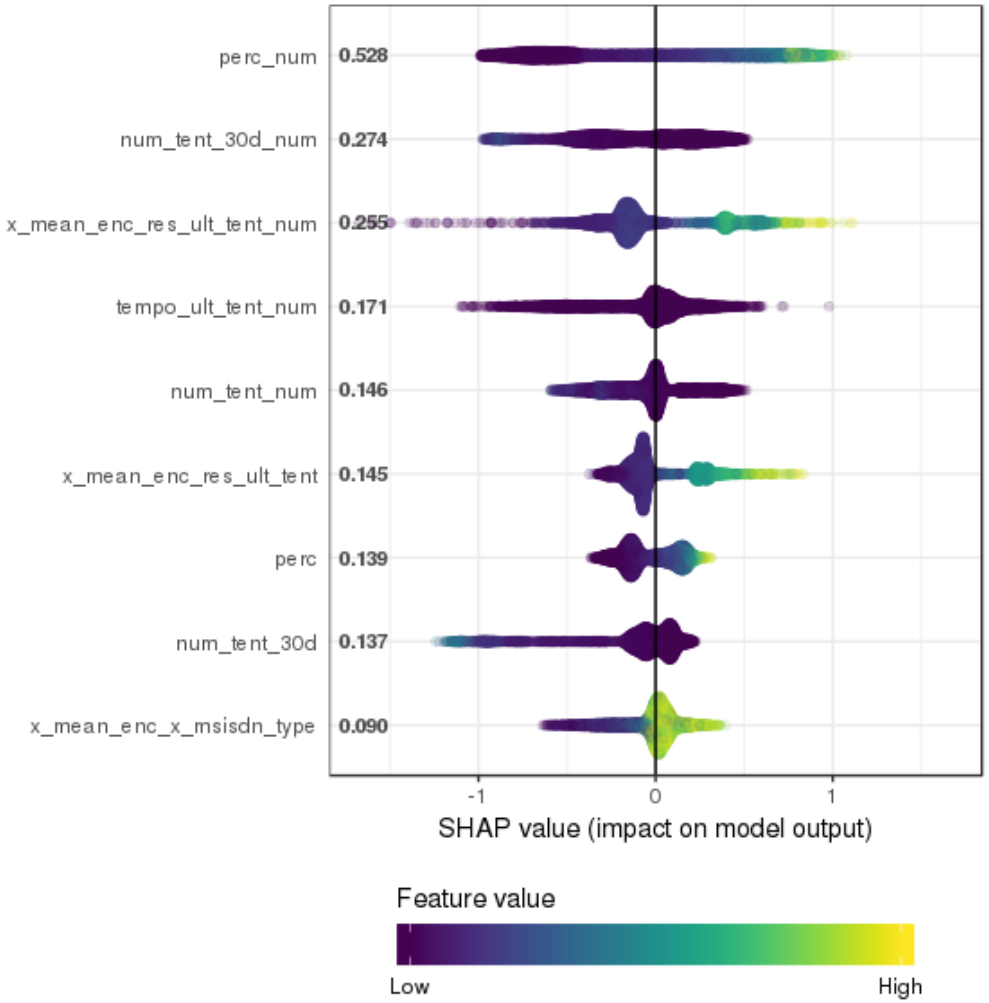
Hour	Value	Mean Target
10	Handled	35%
10	Busy	20%
…	…	…
11	Handled	37%
11	Busy	21%
…	…	…

Predictive Model





Predictive Model

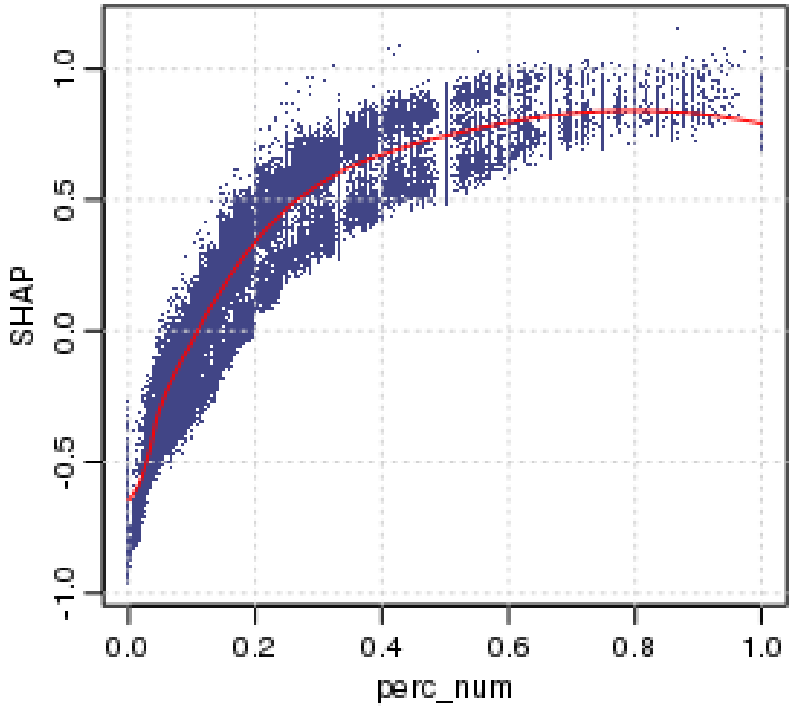


https://github.com/pablo14/shap-values

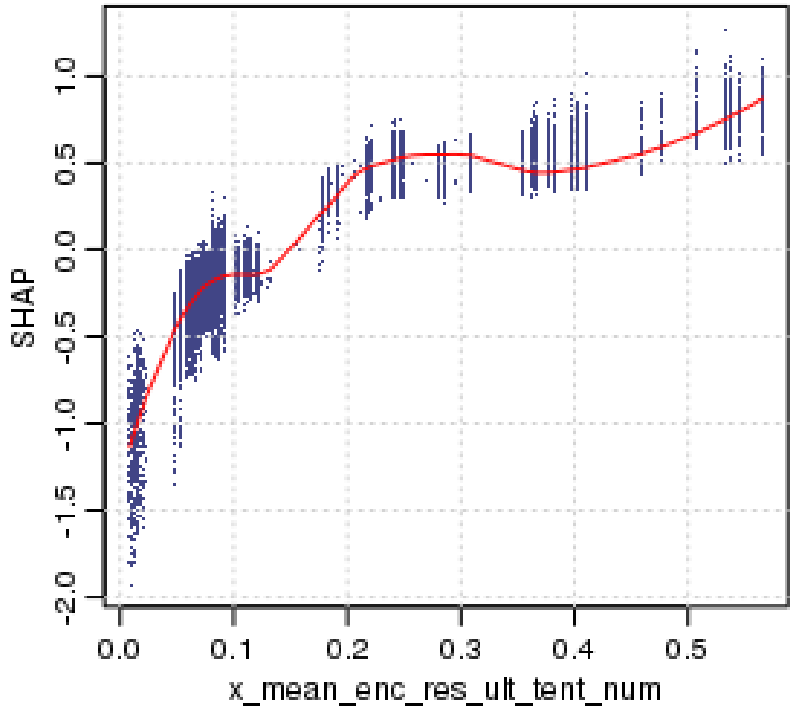
```
shap <- shap.score.rank(model, X_train)
plot_data <- shap.prep(shap, X_train)

plot.shap.summary(plot_data)
xgb.plot.shap(X_train, model)
```

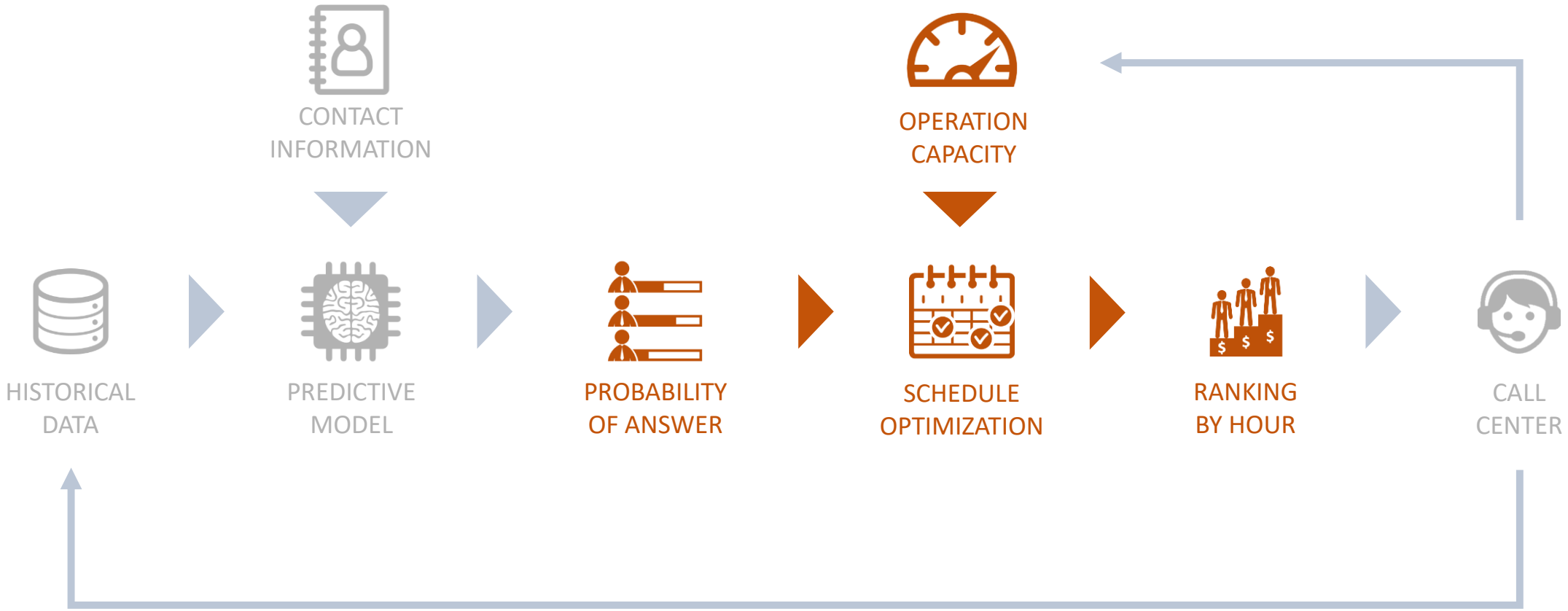
Predictive Model



% of calls made to the same MSISDN at the same hour of day that where answered



Mean Target Encoding of the previous call result (handled, busy,…) made by the same MSISDN


CONTACT INFORMATION
OPERATION CAPACITY
HISTORICAL DATA
PREDICTIVE MODEL
PROBABILITY OF ANSWER
SCHEDULE OPTIMIZATION
RANKING BY HOUR
CALL CENTER

Schedule Optimization

Integer Linear Programming (ILP) can be used to maximize the likelihood of a client answering when subject to the operational constraints.

In practice ILP is computationally slow and requires many assumptions to hold

$$\max_{x} \sum_{n=1}^{N} \sum_{t=1}^{T} P_{n,t} x_{n,t}$$

$$\text{s.t.} \quad x_{n,t} \in \{0,1\}$$

$$\sum_{t=1}^{T} x_{n,t} \leq M, \ \forall_n \in \{1, \dots, N\}$$

$$\sum_{n=1}^{N} x_{n,t} \leq \mathcal{O}, \ \forall_t \in \{1, \dots, T\}$$

$$\sum_{k=t}^{t+\Upsilon} x_{n,k} \leq 1, \ \forall_t \in \{1, \dots, T-\Upsilon\}, \forall_n \in \{1, \dots, N\}$$

"Putting a Call Centre in Peak Performance", IST, João Lourenço, 2018

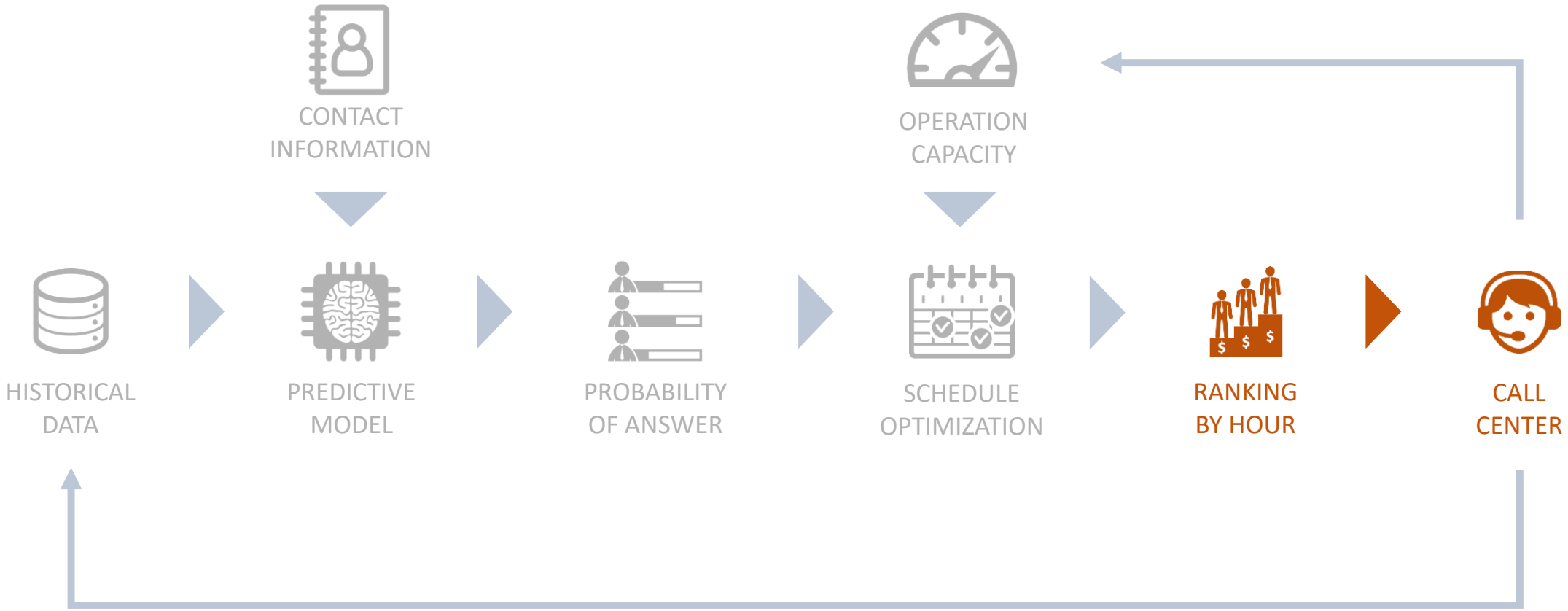
Schedule Optimization

Heuristics allow you to obtain a good enough solution in a fraction of the time.

Minimax Regret (a.k.a Savage Criterion)

Providing a ranking for all MSISDNs instead of a list of MSISDNs to call at each hour, allows for dynamic scalling of operation size.

	10h	11h	12h	13h	14h	15h
MSISDN 1	10,9%	14,2%	15,1%	19,3%	3,8%	15,2%
MSISDN 2	9,8%	5,2%	5,1%	10,6%	14,2%	12,8%
MSISDN 3	1,3%	1,9%	2,0%	2,3%	0,7%	1,0%
MSISDN 4	18,2%	32,1%	25,9%	28,4%	2,4%	26,8%
Opportunity cost	8,4%	5,1%	4,3%	**0%**	15,5%	4,1%
	4,4%	9,0%	9,1%	3,6%	**0%**	1,4%
	1,0%	0,3%	0,2%	**0%**	1,5%	1,3%
	13,9%	**0%**	6,2%	3,7%	29,7%	5,3%
Result	#3	#3	#2	#1	#3	#3
	#2	#4	#4	#3	#1	#2
	#1	#2	#1	#1	#2	#1
	#4	#1	#3	#4	#4	#4


CONTACT INFORMATION
OPERATION CAPACITY
HISTORICAL DATA
PREDICTIVE MODEL
PROBABILITY OF ANSWER
SCHEDULE OPTIMIZATION
RANKING BY HOUR
CALL CENTER

Road to Production

Pilot BTC

Proof of Concept yields +20% Answer Rate

Jul

Technical Debt

Can't reproduce results.
Code has separate paths for feature generation.
Full Code Review.

Aug/Sep

Production Launch

Production launch in one operation.

Oct

Ranking not followed

Dialer doesn't follow ranking correctly due to contract misspecification.

Nov

BTC "100%"

Dialer implementation fixed.

Nov

Troubleshooting

More operational problems – infrastructure, data issues, ranking still not being followed

Nov/Dec

BTC "140%"

Problems fixed and relauch.

Jan

Partner's specialist exit

No ability to scape up operations within that partner.
Expansion on hold.

Feb

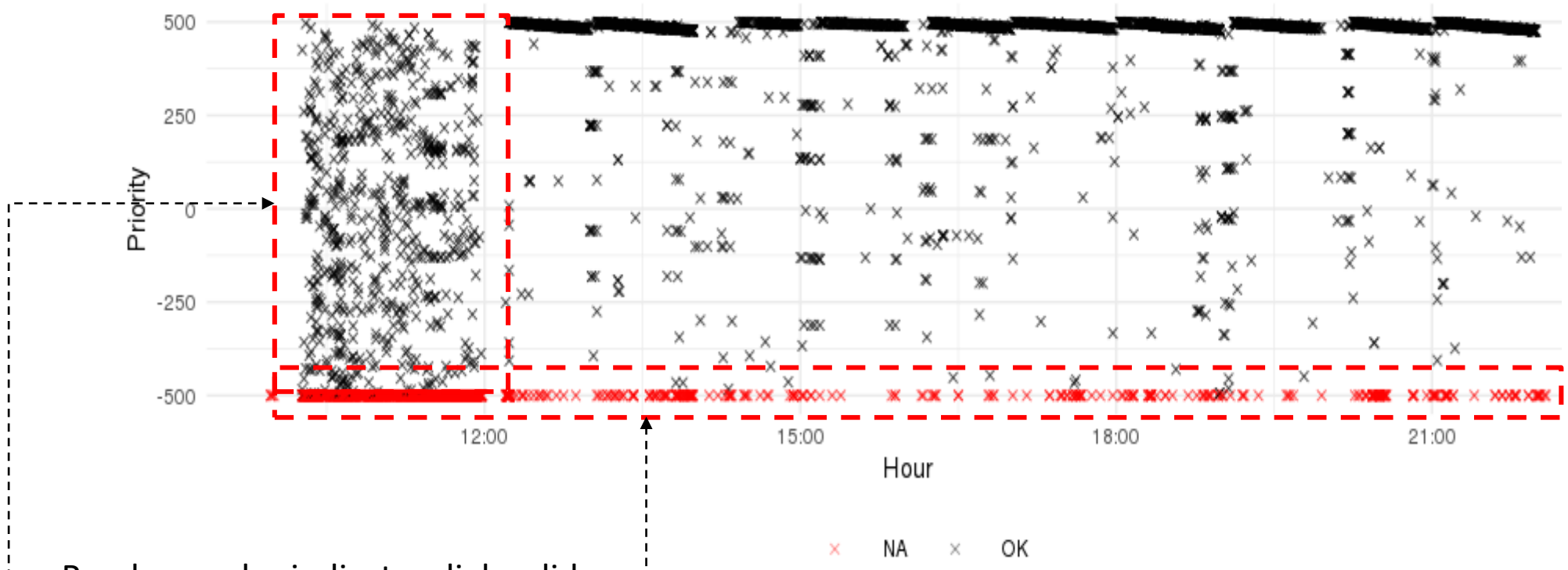
BTC 1.0

Hands-off execution.
Configurations, Reporting

Apr/May

+Detail

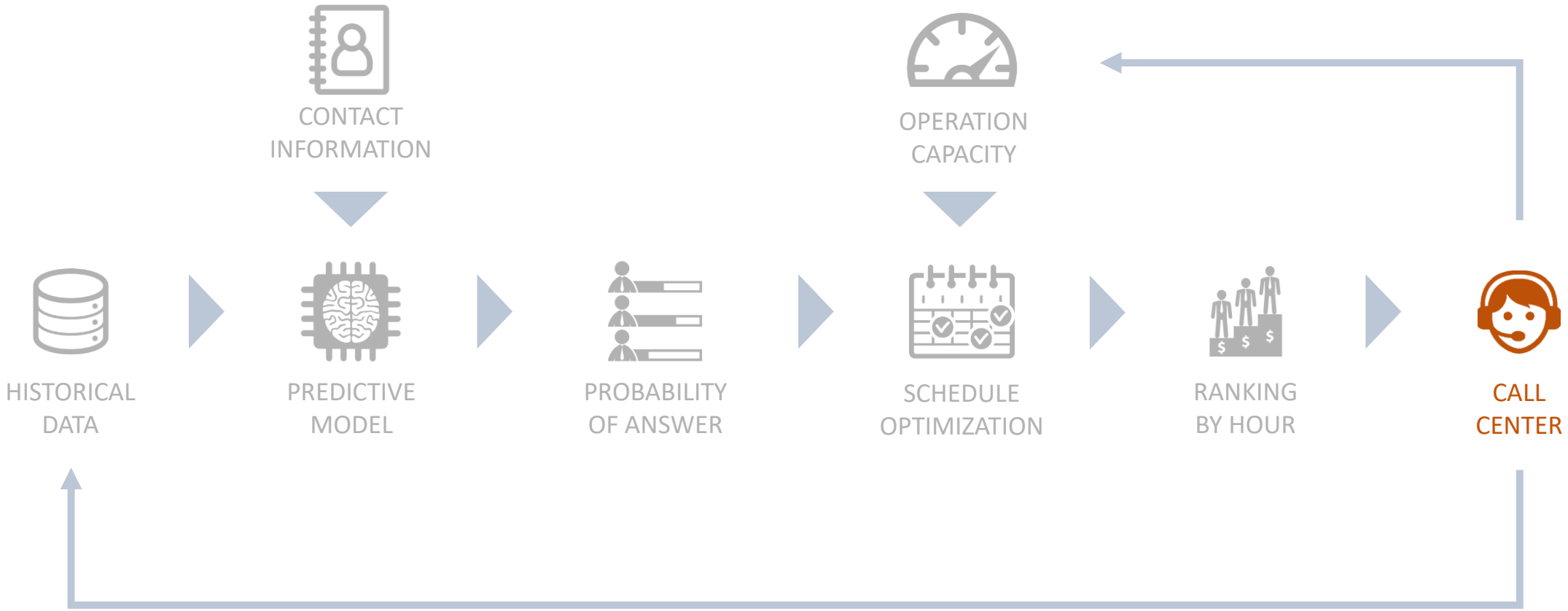
Troubleshooting



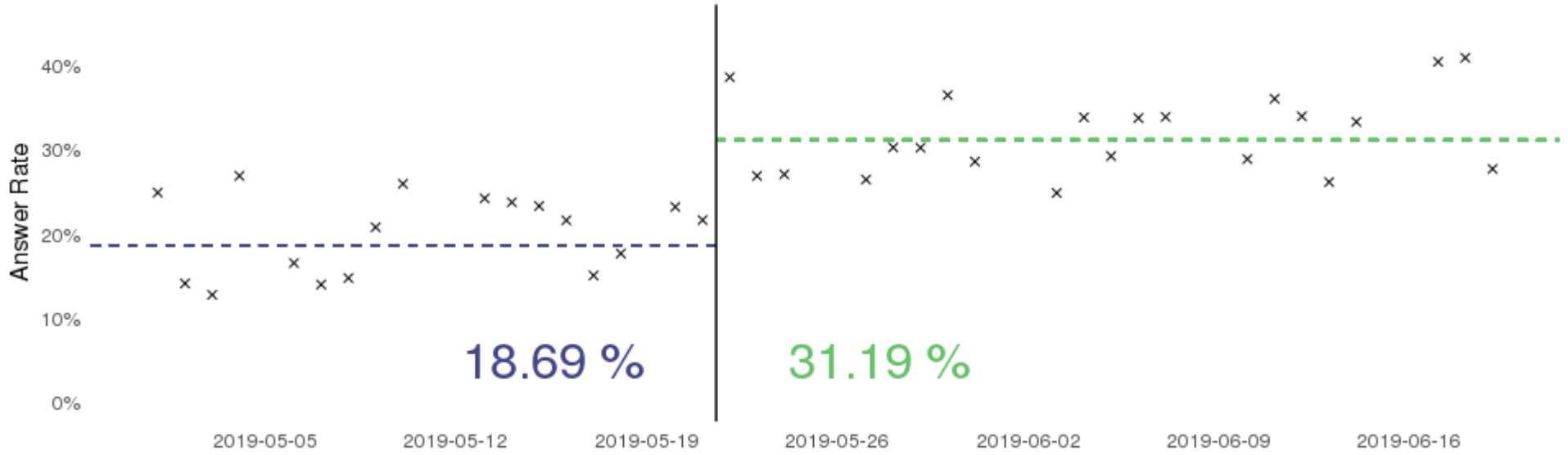
Random order indicates dialer did not use BTC ranking before 12:00

'NAs' are calls to MSISDN's that where not sent in ranking

Data analysis can be used to pinpoint bugs in systems you don't even have access to


CONTACT INFORMATION
OPERATION CAPACITY
HISTORICAL DATA
PREDICTIVE MODEL
PROBABILITY OF ANSWER
SCHEDULE OPTIMIZATION
RANKING BY HOUR
CALL CENTER

Results



67%

Increase in Answer Rate

32% (1)

Decrease in Call Attempts

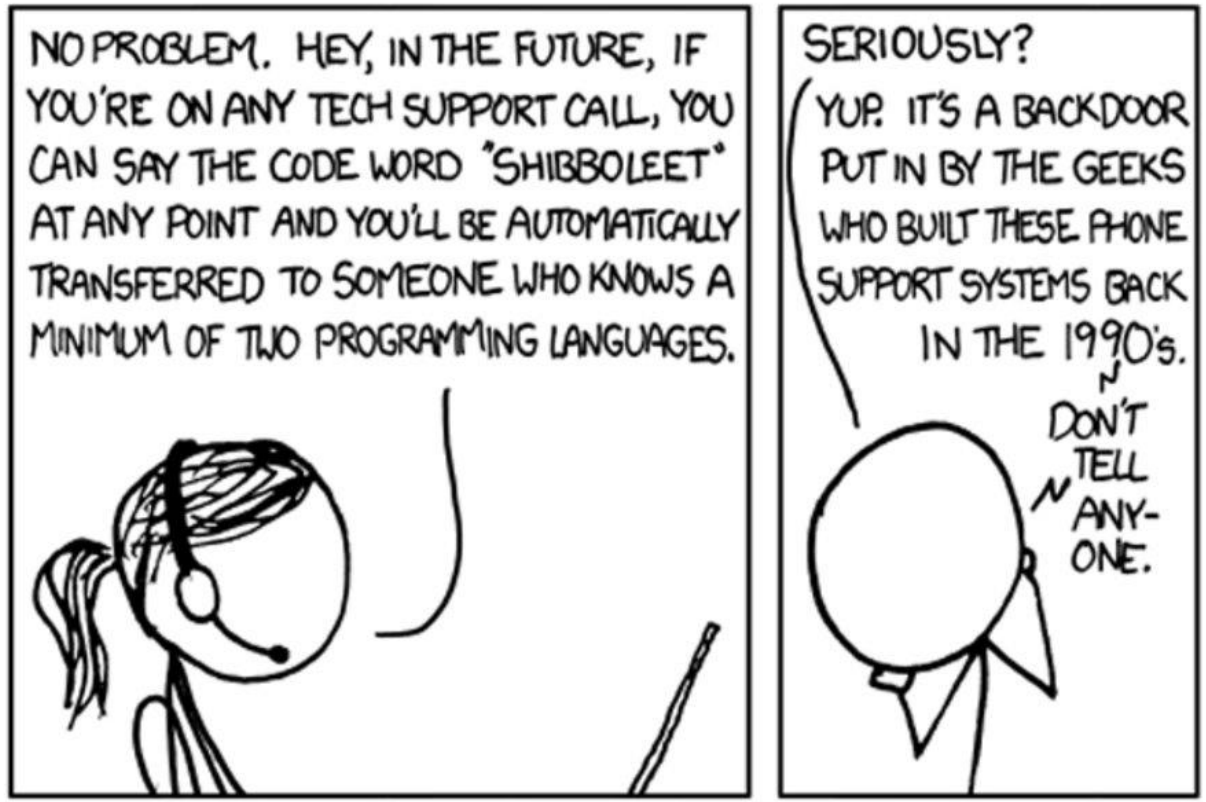
15% (1)

Increase in Client Reach

(1) Results from a previous analysis on a different operation

Future Work

- Scale up
  - Move to fully distributed architecture
  - Scale up to many more operations
- Control Groups
  - Measure BTC impact in the long-run
  - Ensure we can keep learning
- Revisit design choices (scheduling mostly)
- Try out Reinforcement Learning for solving prediction and scheduling in one go
- Improve ops tools (input data validation)
- Incremental stuff (better data cleaning, better features)



Thank You!

UPPER DELTA

NOS