Model Industrialization in ING Bank

Presentation to Data Innovation Summit - 2019

Dor Kedem

2019-03-15

I will not waste your time

You will learn something about:

Data science activities in the banking domain.

Using data science in transforming your organization.

Scaling up machine learning applications in large organizations.

Get the slides - Lots of useful references for follow up

Linkedin.com/in/KedemDor
Dor.Kedem (at) ing.com

Image Credit: My wife, adorageek.com

A bit about me – Dor Kedem

- **Extensive software development career since 2002.**
- **Working on AI research & data science applications since 2010.**
- **At ING Bank in Amsterdam since 2014.**
- **Today, a lead data scientist and product owner.**

ING

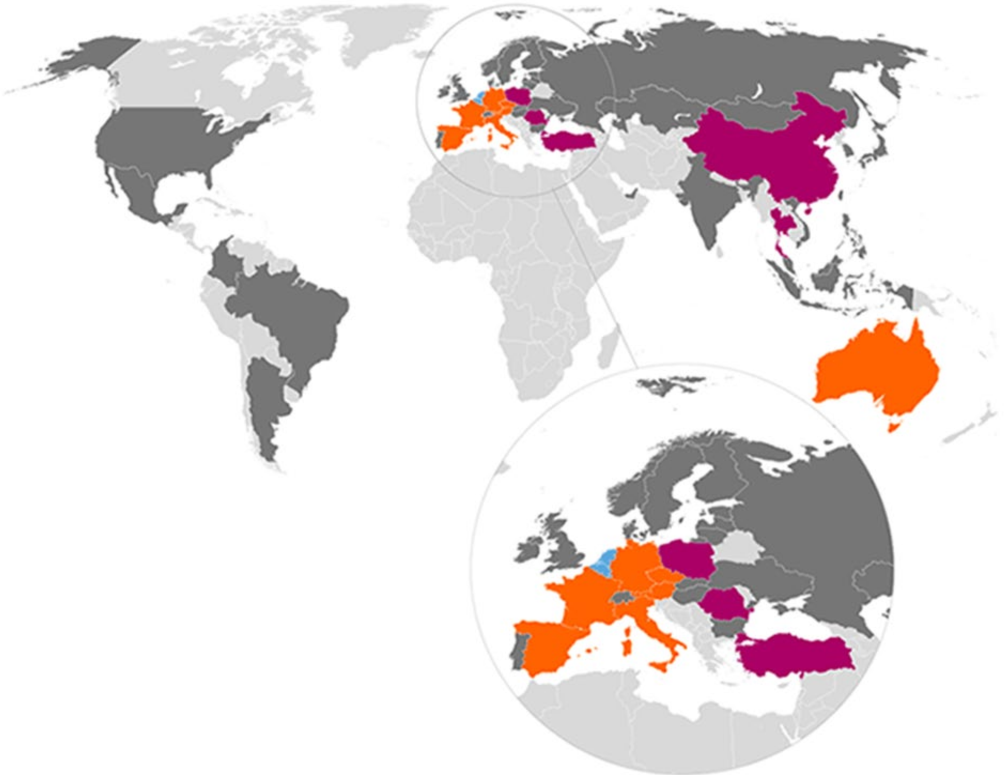
Grab me later (or via LinkedIn) to talk about:

- CI/CD solutions for a data science project lifecycle.
- Impact-driven data science (moving from POCs to MVPs mindset).
- Modelling techniques and machine learning applications in banking.
- Transitioning from software development or IT roles to data science.
- Board games and 3D puzzles.

Linkedin.com/in/KedemDor
Dor.Kedem (at) ing.com

Image Credit: My wife, adorageek.com

ING Bank at a glance



Market Leaders
Netherlands, Belgium, Luxembourg
- Leading retail and wholesale banks in the Benelux
- Evolving into ‘direct-first’ banks
- Improving operational excellence

Challengers
Germany, Austria, Spain, Italy, France, Australia, Czech Rep.
- Leading ‘direct-first’ banks
- Further integrating retail and wholesale banking capabilities
- Broadening lending capabilities

Growth Markets
Poland, Romania, Turkey and our stakes in Asia
- Strong positions in fast-growing countries
- Evolving into ‘direct-first’ banks
- Developing digital leadership capabilities

Wholesale Banking network and global franchises
- International network: more than 40 countries
- Extensive international client base across all regions
- Global franchises: Industry Lending and Financial Markets; Trade Finance and Cash Management

Active in more than 40 countries

38M retail customers and 12.5M primary customers in 4Q18

+54.000 employees in ING Group

Net Promoter Scores: #1 in 6 out of 13 retail countries

Source: https://www.ing.com/About-us/Profile/Key-figures.htm

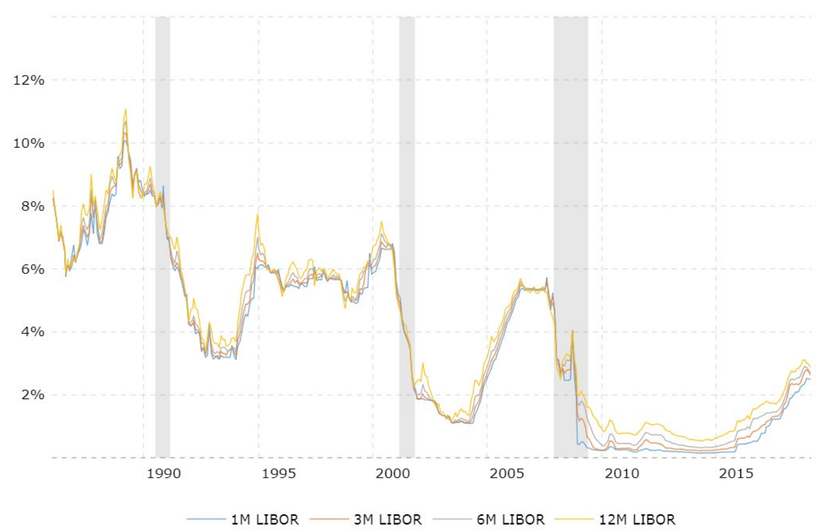
Challenges in European Banking Scene

Historically low interest rates

Regulations leads to more open banking

Fintech is everywhere…

Historical LIBOR rates (grey – recession)



Source: macrotrends.net

Source: https://hollandfintech.com/

How does a bank differentiate itself from the rest?


Forbes Billionaires Innovation Leadership Money

The World's Best Banks: ING And Citibank Lead The Way

Kurt Badenhausen Forbes Staff
SportsMoney
I cover sports business with rare dips into b-schools, local economies

Photo credit: Getty Images GETTY

Dutch financial services giant ING Group has a long legacy of innovation

Our purpose

Empowering people to stay a step ahead in life and in business

Our strategic priorities

Creating a differentiating customer experience

1 Earn the primary relationship

2 Develop analytics skills to understand our customers better

3 Increase the pace of innovation to serve changing customer needs

4 Think beyond traditional banking to develop new services and business models

Sources: https://www.forbes.com/sites/kurtbadenhausen/2019/03/04/the-worlds-best-banks-ing-and-citibank-lead-the-way/ (March 2019)
https://www.ing.com/About-us

Analytics Efforts in ING

Artificial Intelligence: *Currently, ING employs around 80 data scientists, working on various AI-projects:*

Analytics training: *Thousands of employees to engage analytical projects, tools and insights.*


Finextra

ING builds analytics academy to help employees with data skills

18 January 2019

ING is setting up an 'analytics academy' where any member of the Dutch bank's staff can brush up on their data skills.

"Data is the language of the future.

If you don't speak it yet, we'll help you master it."

Görkem Köseoğlu, ING's chief analytics officer.

Source: [Finextra: ING builds analytics academy]

One-to-One Analytics

Our ambition: all customer interactions driven by analytics

Maximising number of analytics driven service and sales interactions

Data > insight > action is in ING's DNA

Democratize big data usage across ING

Users of our services are extremely happy

Data Analytics for customer interactions (NL+BE)

CJE - Christina

DA - Arjen

DS - Samir

DE - Eleanor

	Customer Journey Experts	Data Analysts	Data Scientists	Data Engineers
How many?	Over 400 (outside 1:1)	Over 100	Roughly 20	Roughly 15
What do we know?	• Banking • Marketing theory • Customer engagement • Message framing	• BI tools (SAS, IBM Cognos) • Data Privacy • SQL	• Statistics & ML • Data Privacy • Programming (i.e. Python, R, Scala)	• Big data technologies • CI/CD solutions • Security & Compliance
What do we create?	• Product specification • Online & offline content • Customer engagement	• Reports • Dashboards • A/B Testing	• Statistical models • Data Products	• ETL systems • Data lake • Model hosting

The need for model industrialization

Example case: Credit Card Acquisition

For Black-Friday (Nov 23rd, 2018), Christina wants to contact customers to acquire a new credit card (via website offering or direct communication).
We have two types of offers: regular credit cards & platinum credit card.

How can she find who to contact with these offerings?

CJE - Christina

DA - Arjen

DS - Samir

- Plot customer engagements on different demographics.
- Come up with business rules based on shared personal understanding.

- Build a likelihood model based on past behavior and engagements.
- Rank customers according to this model.

ING

Example case: Credit Card Acquisition

Before Black-Friday (Nov 23rd, 2018), Christina wants to contact customers to acquire a new credit card (via website offering or direct communication).
We have two types of offers: regular credit cards & platinum credit card.

How can she find who to contact with these offerings?

CJE - Christina

Very vast majority

DA - Arjen

- Plot customer engagements on different demographics.
- Come up with business rules based on shared personal understanding.

DS - Samir

- Build a likelihood model based on past behavior and engagements.
- Rank customers according to this model.

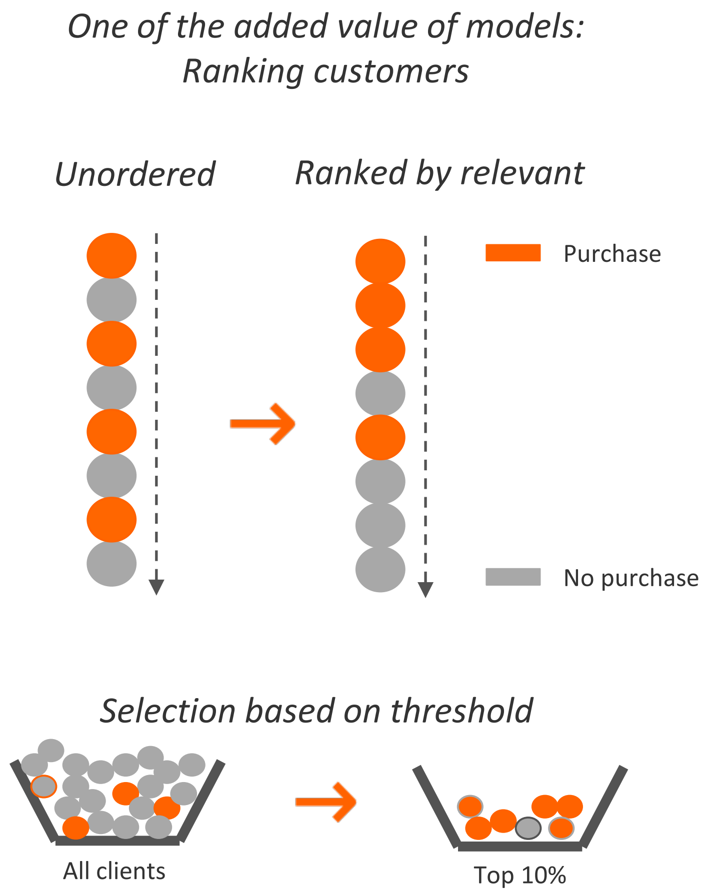
What are we missing when we don’t use models?

It takes a lot of time to make and adjust customer selection.

We’re bound by our personal understanding and our data analyst capabilities.

There’s no structured way of learning and improving our engagements for the next time.

We’re not as relevant or personal to our customers as they expect us to be.



Our Objective

Democratizing model building: Enabling DA's to create models for finding customers for their offers.

Accelerate best practices: Make it easy & fast to be effective in customer selections.

- Model building process "built-in": Tell us "what" you want – we take care of the "how".
- Evaluation "built-in": Decide what to build → Get a free model & campaign evaluation!
- Compliance "built-in": GDPR, archiving, legal, commercial pressure, risk – we got you covered.

CJE - Christina

Understand the customer better

DS - Samir

Saves time
Better engagement

Customer - Claire

More relevant offerings

DA - Arjen

Saves time
Grows in skills

DE - Eleanor

Making large-scale impact

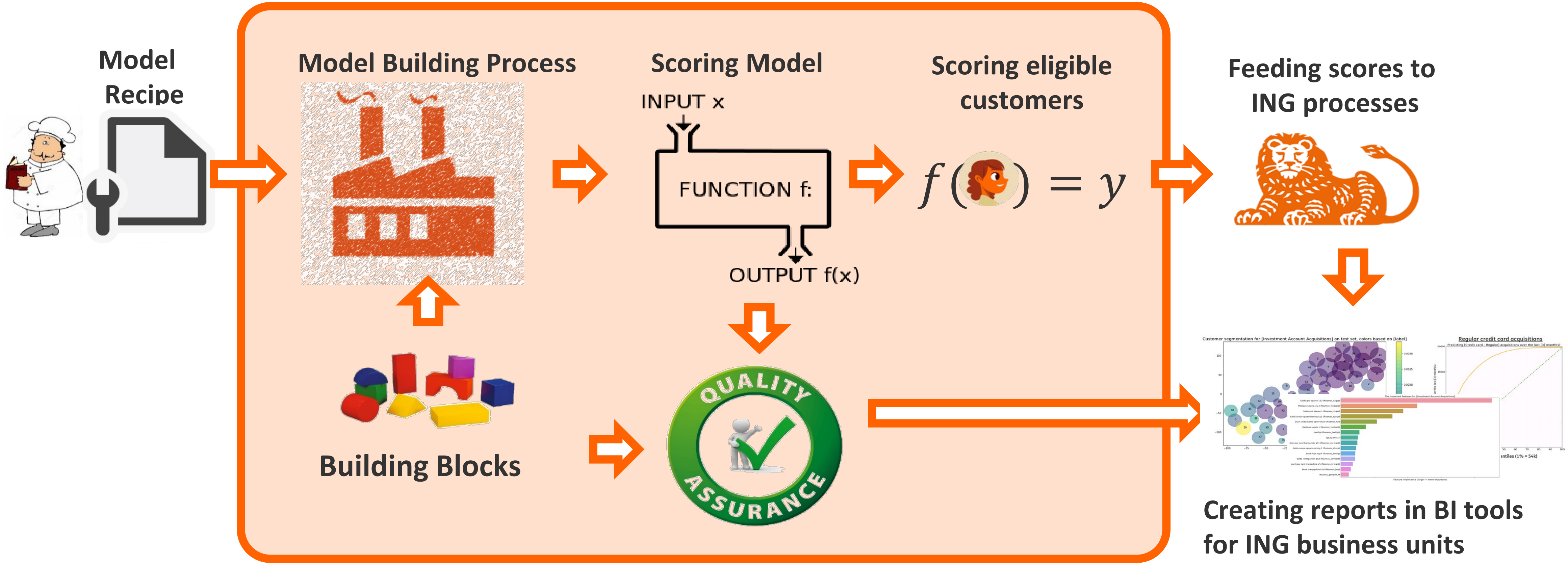
ING Bank

Meeting objectives

Our approach – Model Factory

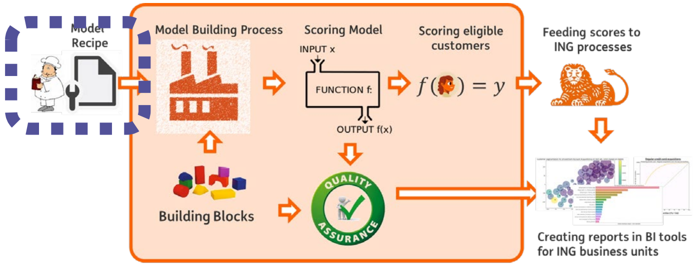
Model Factory

Building customer models without reinventing the model building process



Somewhat similar open source approach: [Uber's Ludwig: Training models without writing any code (February 2019)]

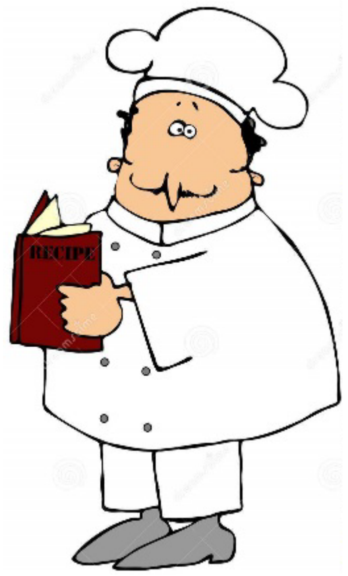
Another open source model factory for reference: [KPN's model factory]



Model Recipe

Mandatory ingredients:

- **Business Objective**
  Selection from: acquisition, deepsell, retention, customer journey.
- **Business Objective specification**
  Based on the objective. For example: which product to acquire?
- **Features to include / exclude**
  Selection from a list. Done based on domain expertise.
- **Customers to include / exclude**
  SQL “where clause”. Based on domain expertise.

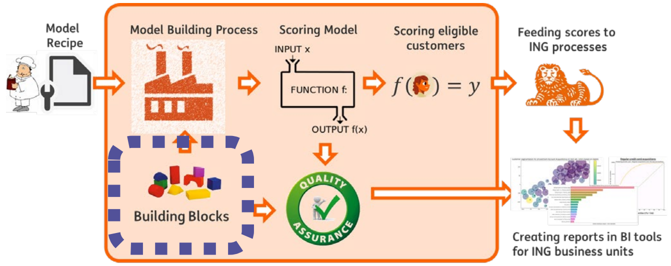


Optional ingredients (with defaults):

- **Times specification:** (How long does it take to acquire, how long before customer makes decision)
- **Modelling techniques:** (for advanced / data scientists users)

Model specification is translated to a 10-15 lines JSON file and is filled by a DA

Building Blocks



Available to all models built with a recipe specification:

Analytics features extraction

Machine learning monitoring processes

Target templates (i.e. acquisition, deepsell)

Classifiers

Evaluators

Hyperparameter / model selection (AutoML)

Fairness & bias reduction

Data-sets creators

Uplift measurement

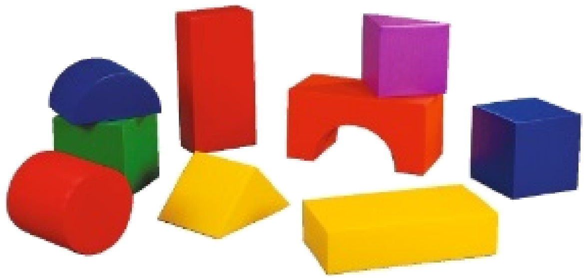
Storage management

Scheduling

Hosting

GDPR applications

Interaction with ING services

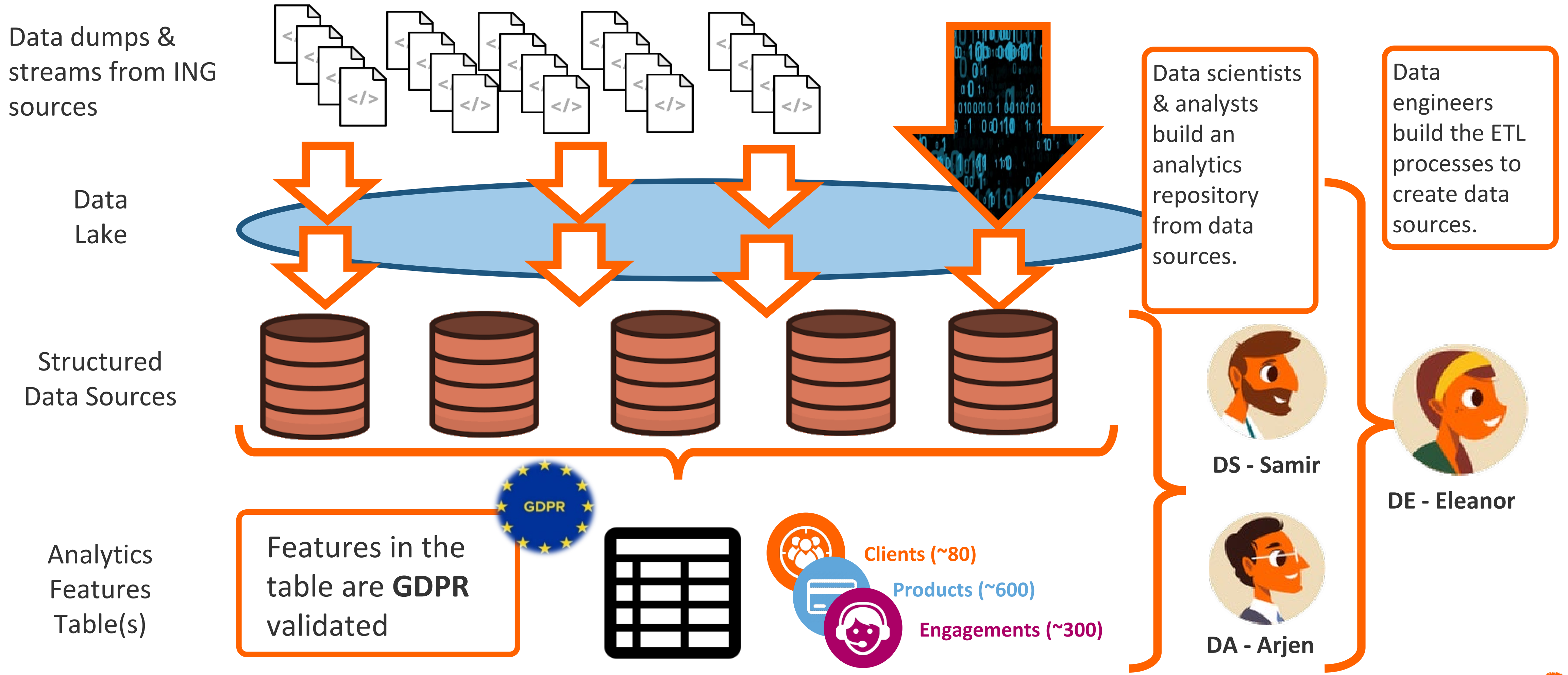


Building Blocks Example (1): Data Sources

Creating the model feature sources

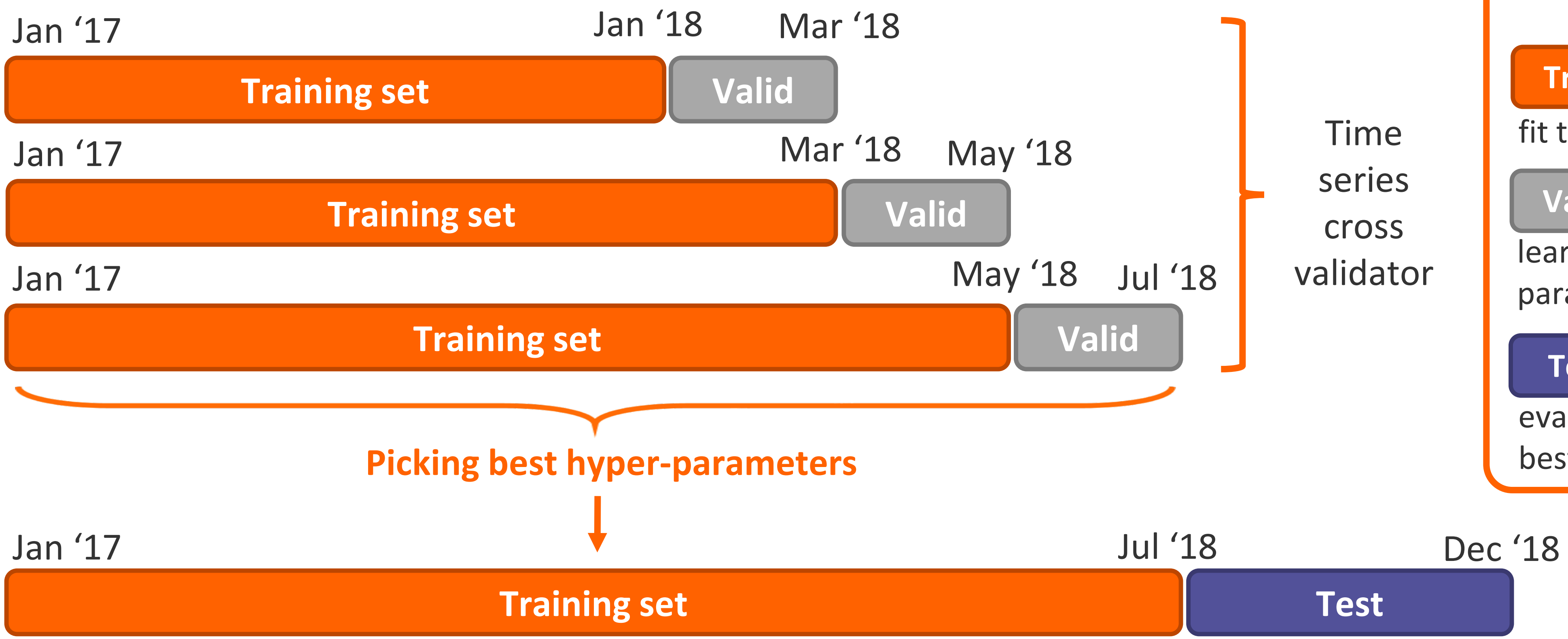
Built on top of:

- IBM PureData for Analytics (PDA)
- SAS Enterprise Global



ING

Building Blocks Example (2): Data Sets Creators



Legend

Train Used to fit the model.

Valid Used to learn hyper-parameters.

Test Used to evaluate and to pick best mode.

<u>Some tips to building datasets:</u>

- Selecting different customers in each timestamps → Generalizing to new customers.
- Arranging data set in time series accordance → Generalizing better for forecasting.

Useful resource - Timothy Lin's [Creating a Custom Cross-Validation Function in PySpark]()

Building Blocks Example (3): Model Building

Relying on open-source Big Data technologies as building blocks

Classifiers (the model types): mainly based on the Spark Machine Learning framework and includes:

- Linear / Logistic regression
- Naïve Bayes
- Decision Trees
- Ensemble methods (Random Forest, GBRT)
- Neural Networks (MLP)

Evaluators (the model performance validation):

- Everything under the Spark MLLib evaluation metrics.

Meta-learning and AutoML (finding the best model):

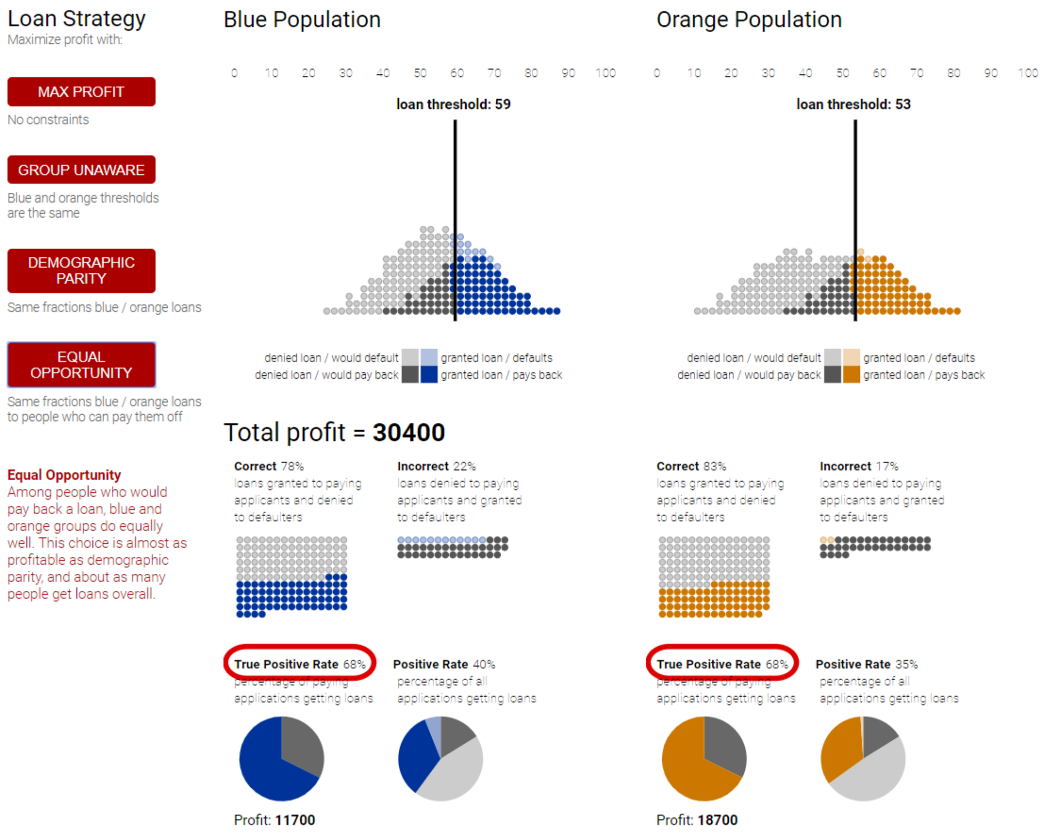
- Currently experimenting with auto-sklearn & H2O for faster hyper-parameter tuning. See Georgian Partners' comparison.

Building Blocks Example (4): Fairness

For easy explanation:

Attacking discrimination with smarter machine learning

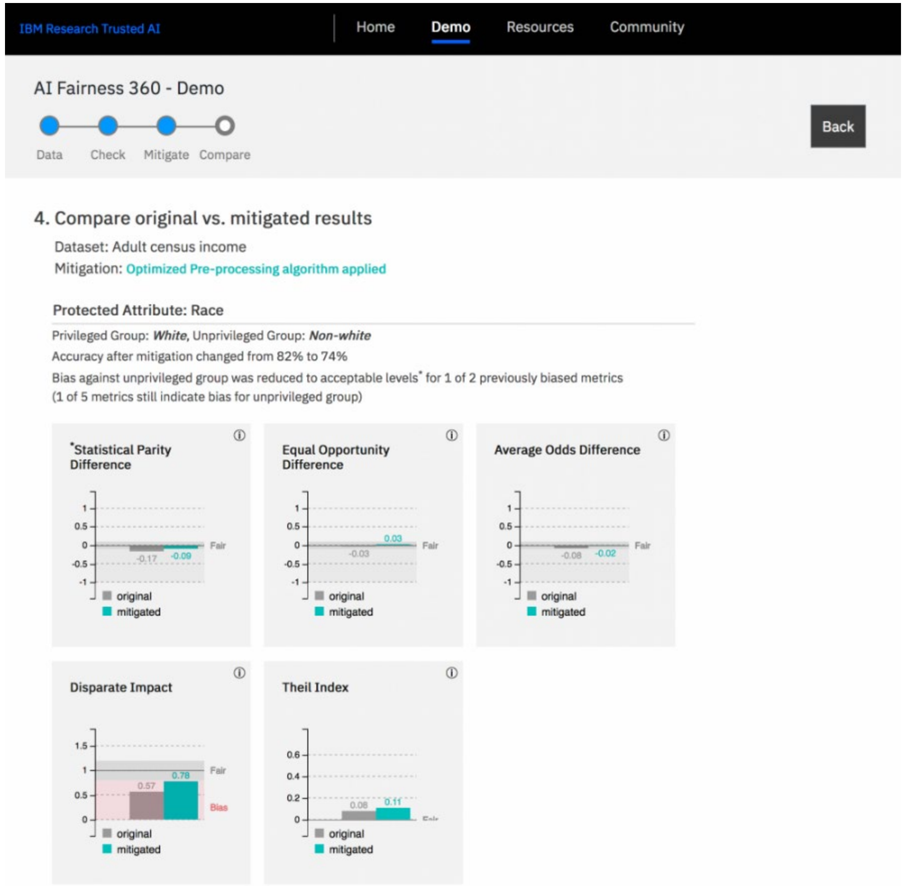
Resource: https://research.google.com/bigpicture/attacking-discrimination-in-ml/



For approaches on reducing bias:

IBM AI Fairness 360

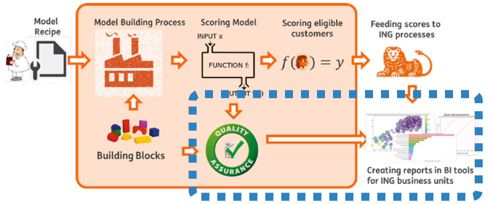
Resource: http://aif360.mybluemix.net/



ING

Model Factory Products

Engaging with the model factory process & results

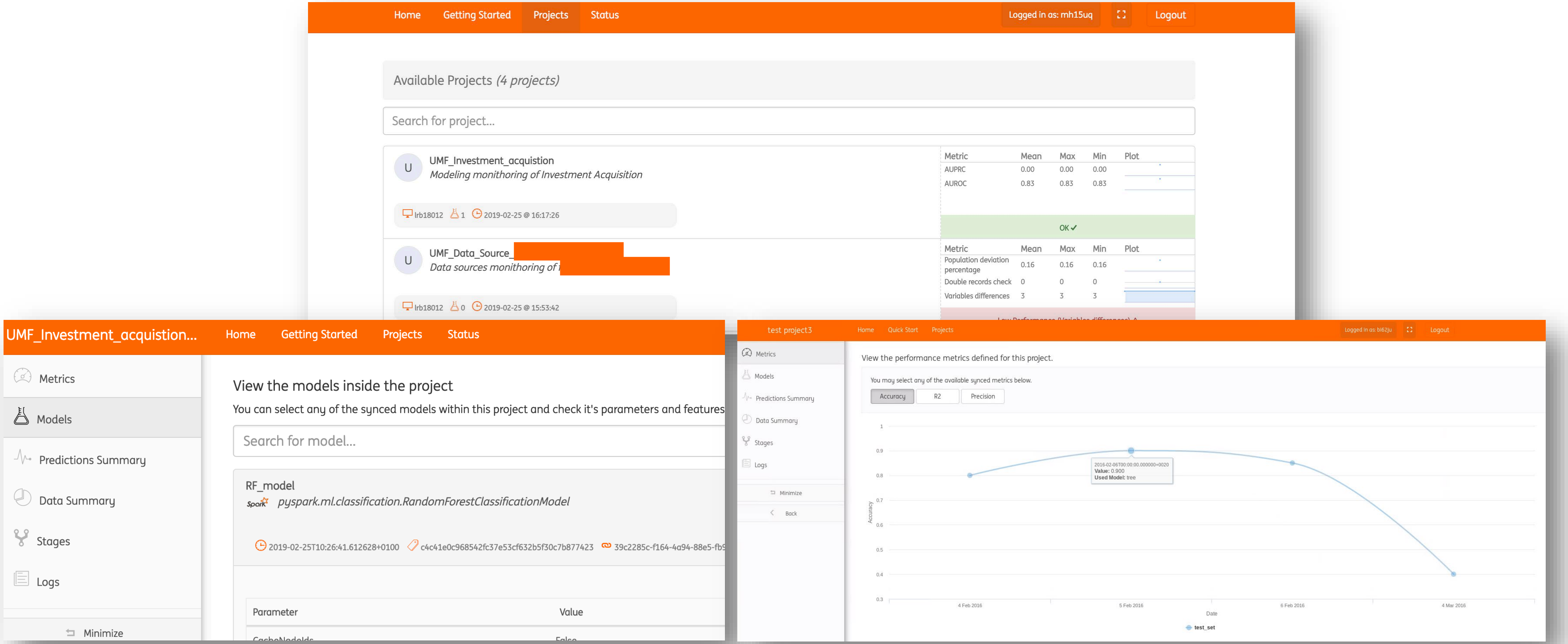


	Validating building blocks	Validating model execution	Validate model quality	Understand the customer better	Selecting customers for campaign	Post-hoc campaign evaluation	Getting the big picture of model usage
Customer Journey Expert				✓	✓	✓	✓
Data Analyst			✓	✓	✓	✓	✓
Data Scientist	✓	✓	✓	✓			✓
Data Engineer	✓	✓					✓

Monitoring Tool (Developed in-house)

BI Tools (IBM Cognos Analytics)

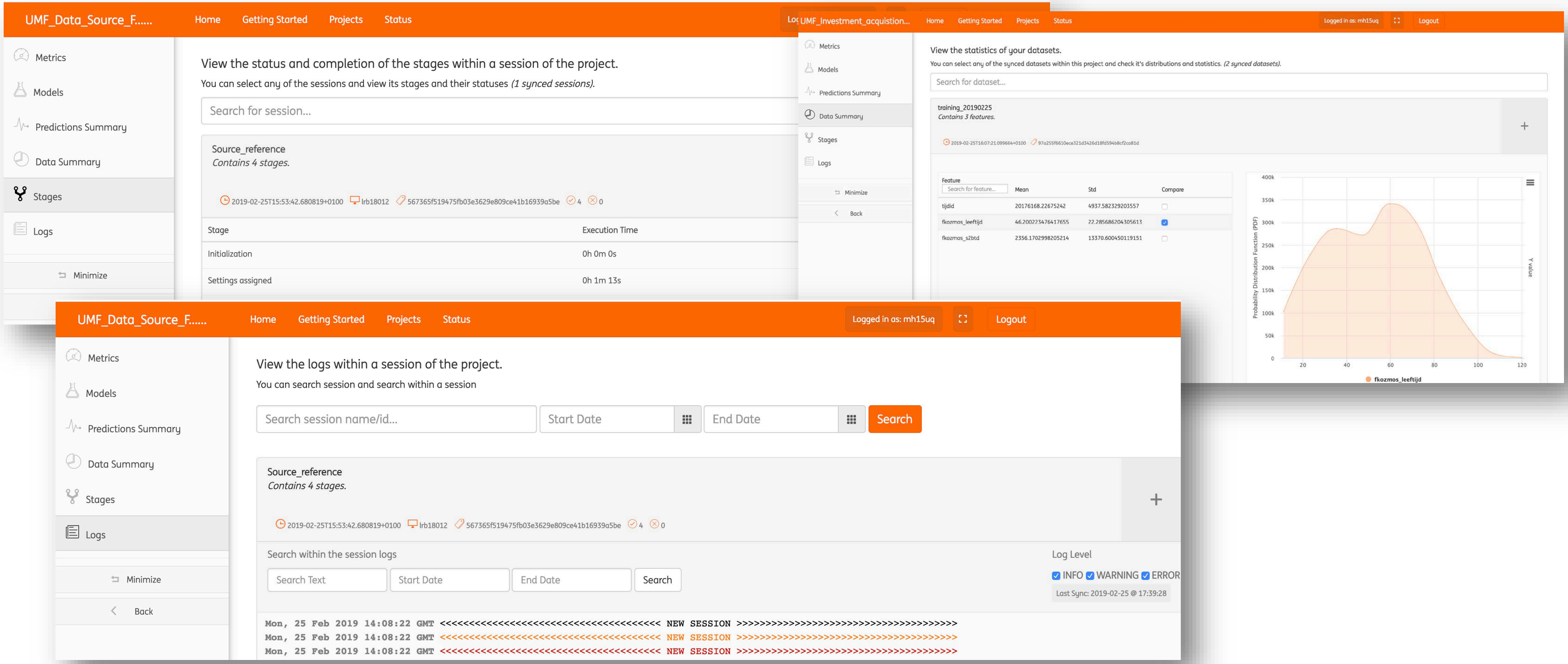
Designated system for monitoring production ML models



Open source alternative: mlflow.org (platform for machine learning lifecycle)

Useful resource: Google AI's [What's your ML test score? A rubric for ML production systems](#) (Breck et. al, 2016)

Designated system for monitoring production ML models



Open source alternative: mlflow.org (platform for machine learning lifecycle),

Useful resource: Google AI's [What's your ML test score? A rubric for ML production systems]() (Breck et. al, 2016)

Reporting on the model built

A. Technical quality metrics

B. Lift curve

C. Cumulative Gains

D. Overlap with manual selection

E. Feature Importance

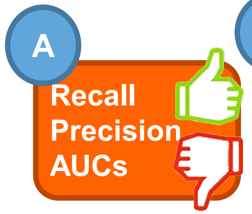
F. Customer Segmentation

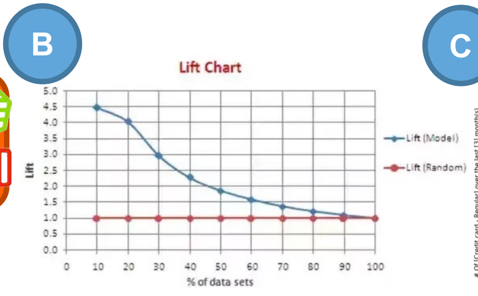
G. Model comparison heat map.

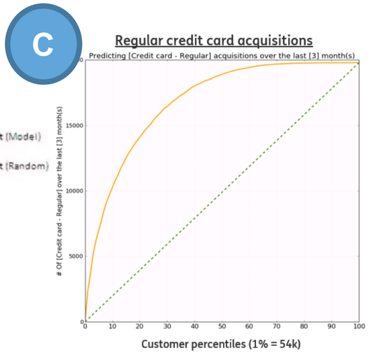
H. Compare features distributions.

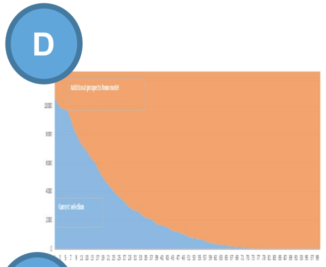
I. Score distribution

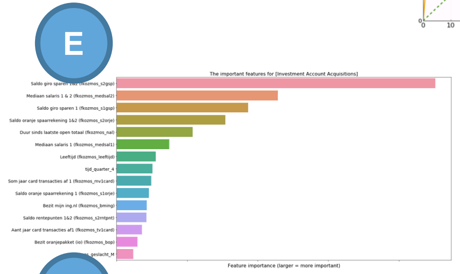
J. Conversion for feature values.

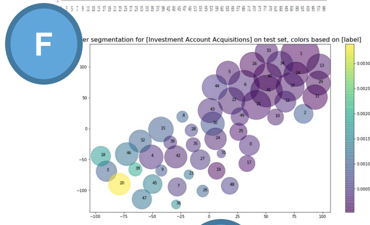


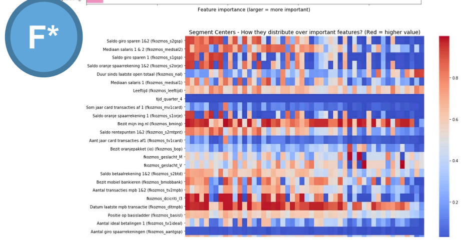


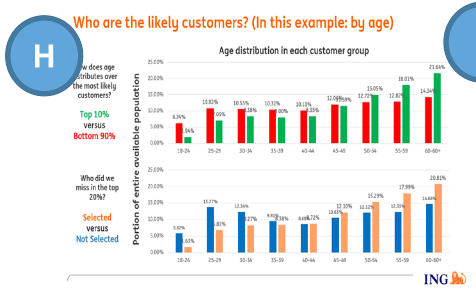


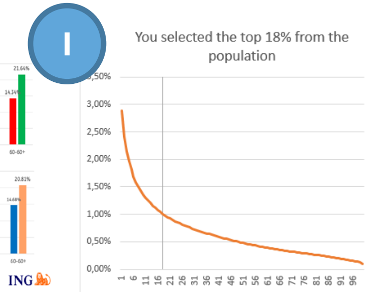


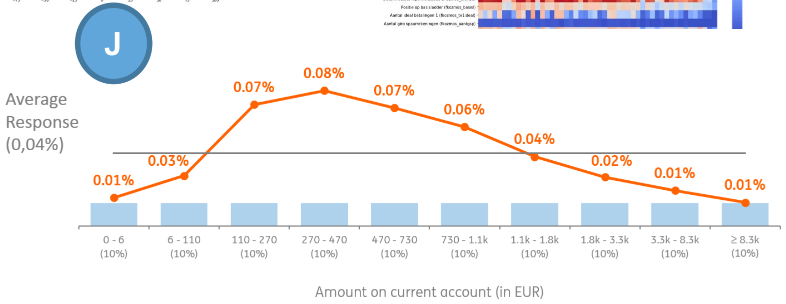






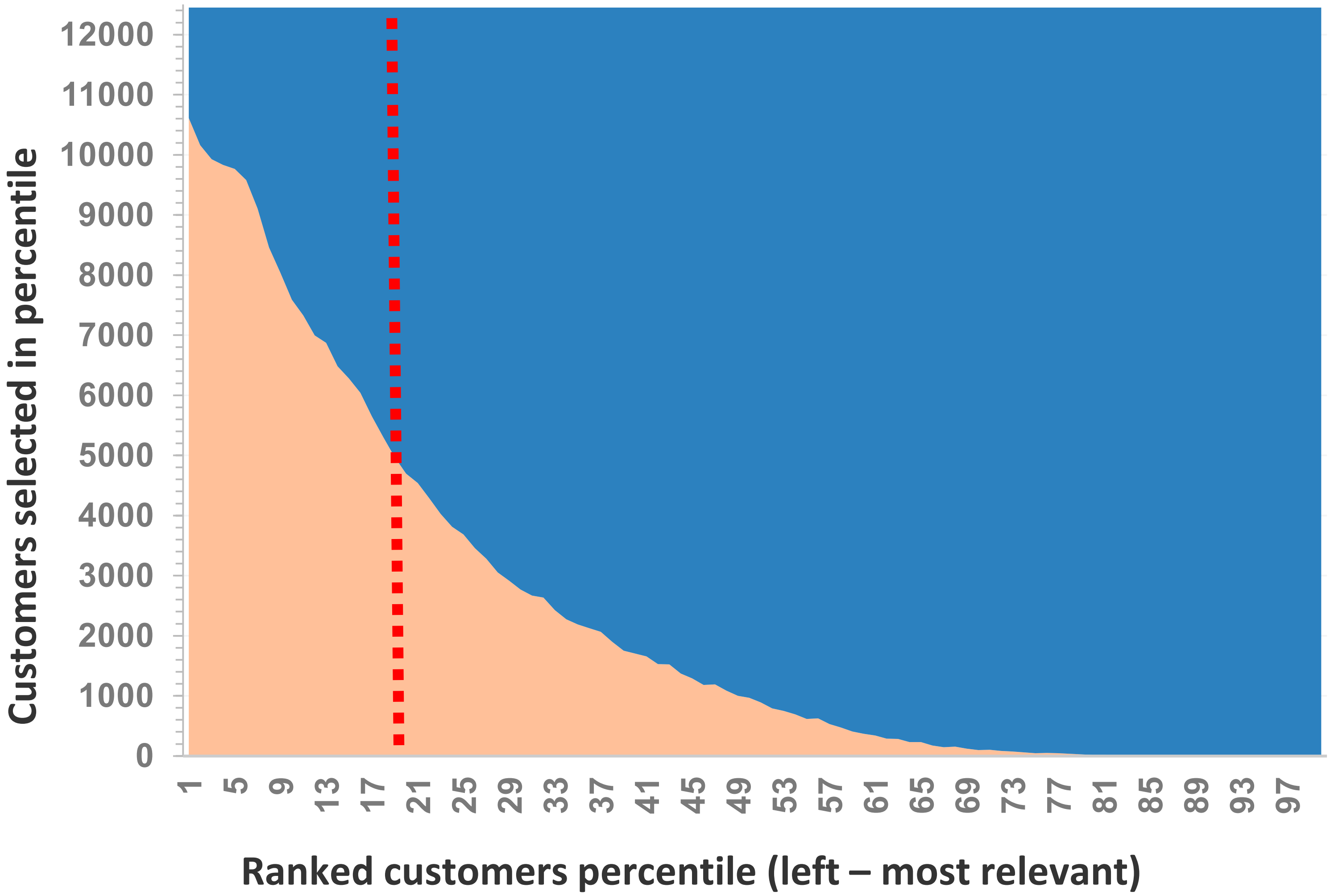






What's the difference between my old selection and the model's?

D



"I feel more confident the model makes meaningful selections"

"I see that the model found top customers that I haven't contacted yet"

DA - Arjen

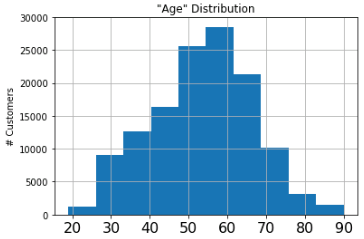
ING

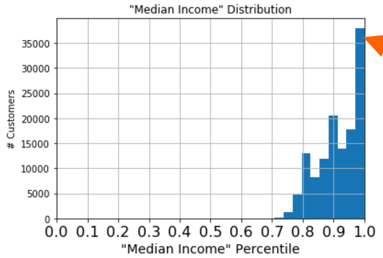
Customer Segmentation

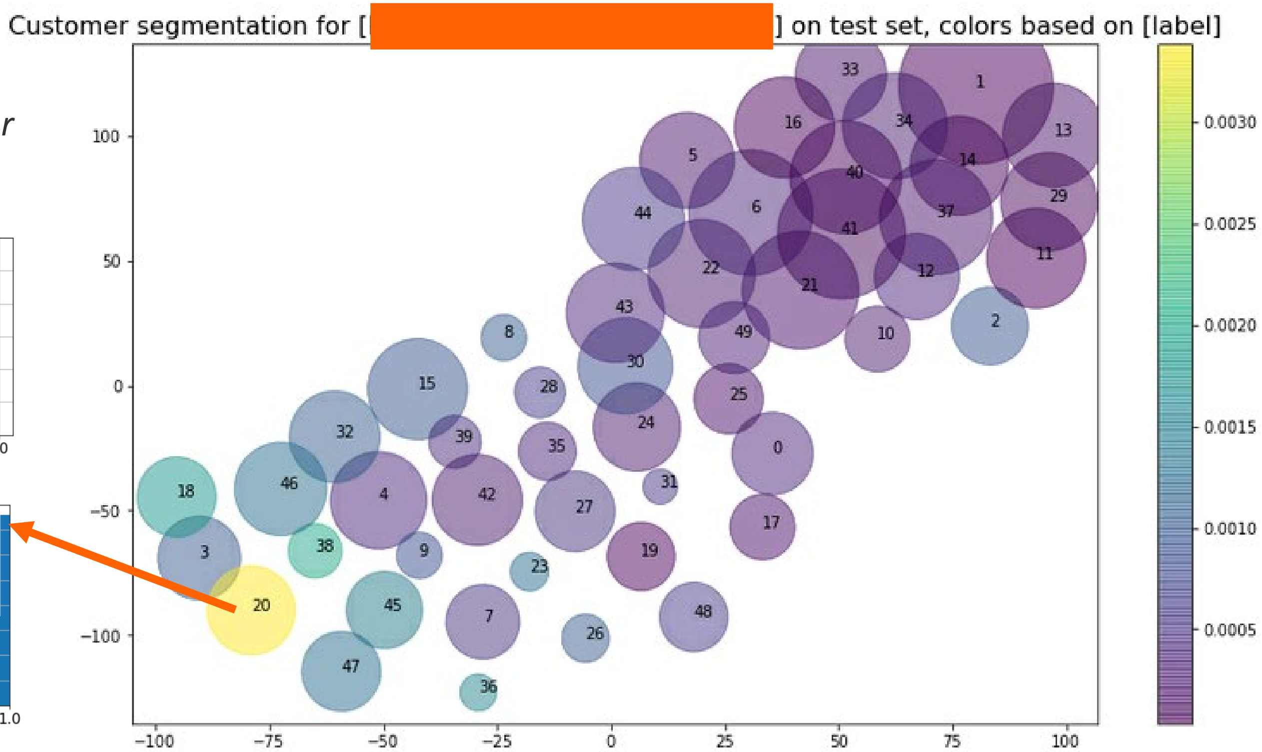
F

Grouping customers together based on the model's important features.

Allows for further analysis on customer segments







Segment size: indication of number of customers.

Segment color: average conversion (more yellow = higher conversion).

Y,X Axes: Don't mean much, but the overall distance between segments mean that customers are more different based on important features (closer segments = more similar).

CJE
Christina

This helps me understand who are my customers and to tailor a message for each type of customers.

ING

Credit Card Acquisition – Which proposal to who?

G

X-axis: Ranked customers interested in regular credit card (left - most interested)
Y-axis: Ranked customers interested in platinum credit cards (down - most interested).

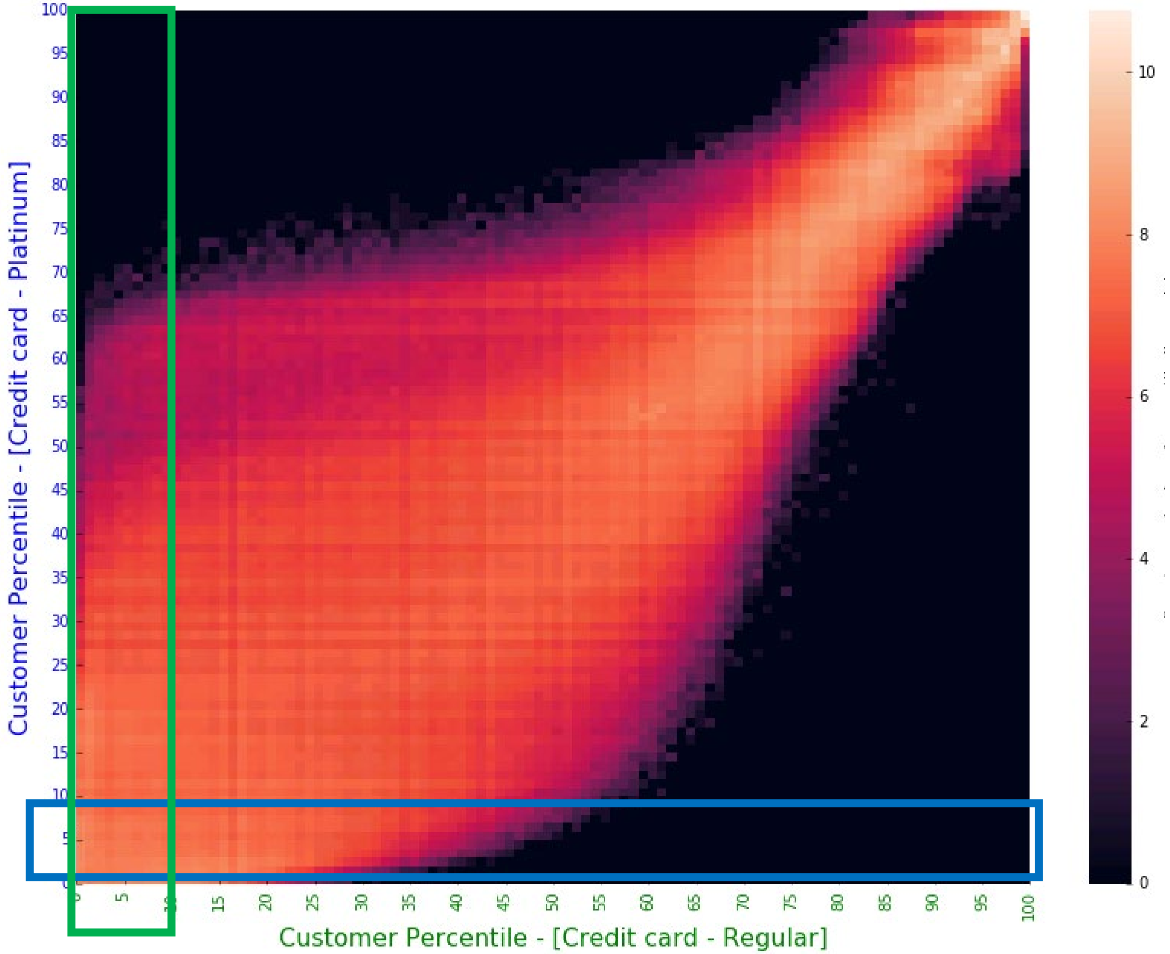
Rectangles – the top 10% of customers in each group.

Bottom 90% Platinum	**347k**	**4.9Mil**
Top 10% - Platinum	**232k**	**412k**
	Top 10% Regular	**Bottom 90% Regular**

DA - Arjen

"I can now send the relevant offer to the relevant customers and avoid spamming.

Combined ranking for both credit card acquisition models



customers in shared percentile (log scale)

Brighter = more customers

ING

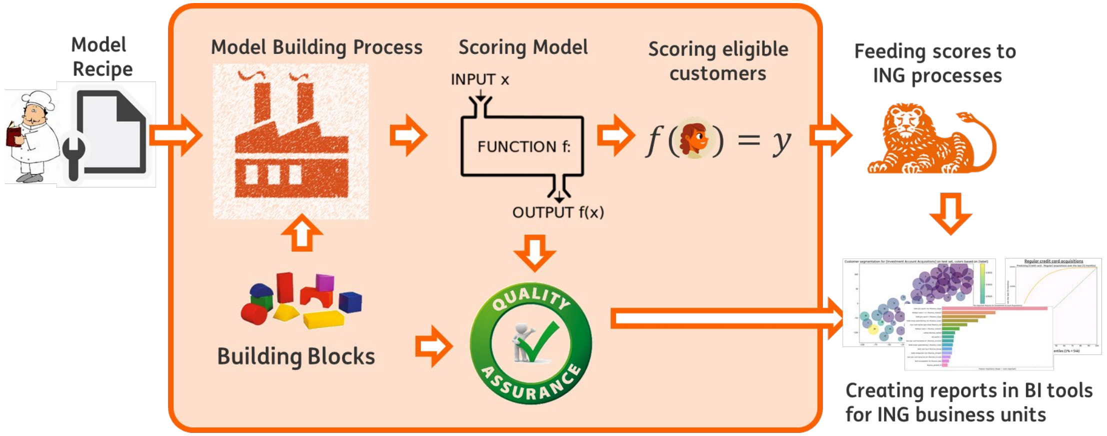
To wrap up

Model Industrialization in ING Bank

Linkedin.com/in/KedemDor
Dor.Kedem (at) ing.com

Summary:

- Enabling model creation, without coding and using data scientists best practices and cumulative efforts.
- Simple specification, modular design.
- Accelerates DA's, empowers CJE's, and makes all of us more relevant to our customers.



Selected Resources:

Driving innovation:

- ING PACE: Evidence-based design-driven lean approach

Model building:

- Uber's Ludwig – Building models without coding
- Georgian Partners' AutoML comparison
- Creating a Custom Cross-Validation Function in PySpark
- Distributed deep learning on spark: dist-keras

Machine learning in production:

- What's your ML test score? A rubric for ML production systems
- MLFlow: machine learning lifecycle

Fairness & bias removal:

- Google's "Attacking Discrimination in ML"
- IBM's AI Fairness 360