DreamQuark

Sharpen decisions in Financial Services
with High-Scale Production & Trusted AI

Deep learning from physics to financial services

Jeremie Abiteboul

AMLD – AI & Physics Track

January 28, 2020

Who are we ?

Where do we come from ?

Université Paris 7 Denis Diderot
UFR de Physique

Thèse en vue de l'obtention du diplôme de
docteur de l'université de Paris 7 en physique
Ecole doctorale 517, Particules-Noyaux-Cosmos

Etude théorique et expérimentale des corrections électrofaibles au processus de production inclusive de jets
Développement de méthodes de détection de topologies extrêmes

Nicolas MERIC

Thèse réalisée au CEA, LPNHE, LPTHE
Sous la direction de :
Gavin SALAM et Philippe SCHWEMLING

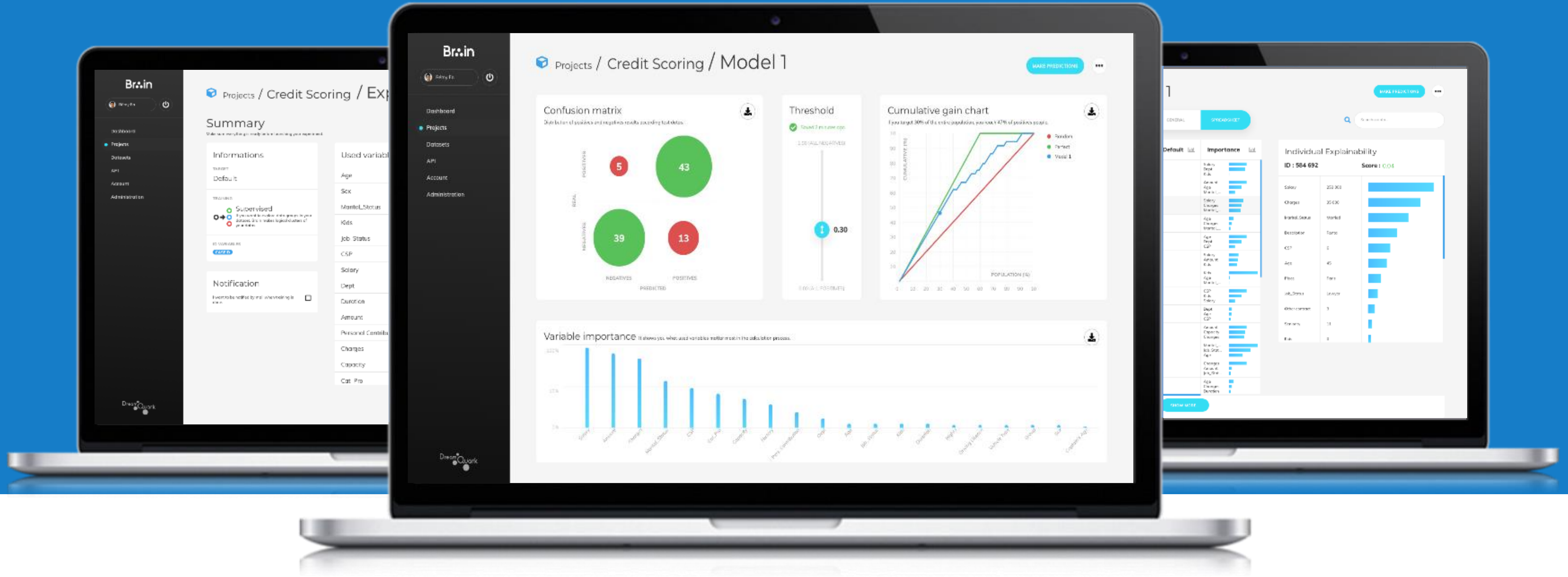
Thèse soutenue le jeudi 19 septembre 2013.

- Deep learning @ATLAS
- DreamQuark established 2014

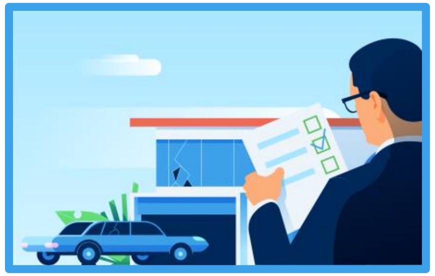
Empower your action thanks to science

- My background: computational plasma physics, turbulence modeling

=> originally applied ML to analyze experimental data & simulation data



BANKS & INSURANCE



Fraud & Compliance

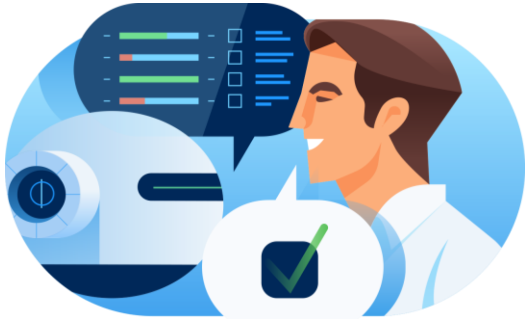
Fraud detection
AML
Compliance supervision

Marketing

Product upsell/cross-sell
Propensity
Customer segmentation
Satisfaction & anti-churn
Product recommendation

Risk

Risk model
Credit scoring
Fragile clients
Recovery



Transparency of decisions

Ability to do more business

Ease of use

Ease of deployment

DreamQuark software is market-proven

THEY TRUST US

RBS The Royal Bank of Scotland

NatWest

BNP PARIBAS

bpifrance

gan

vaudoise

AXA

AG2R LA MONDIALE

RECENT AWARDS

10 000 STARTUPS POUR CHANGER LE MONDE LAURÉAT 2018

Pay FORUM AWARDS 2018

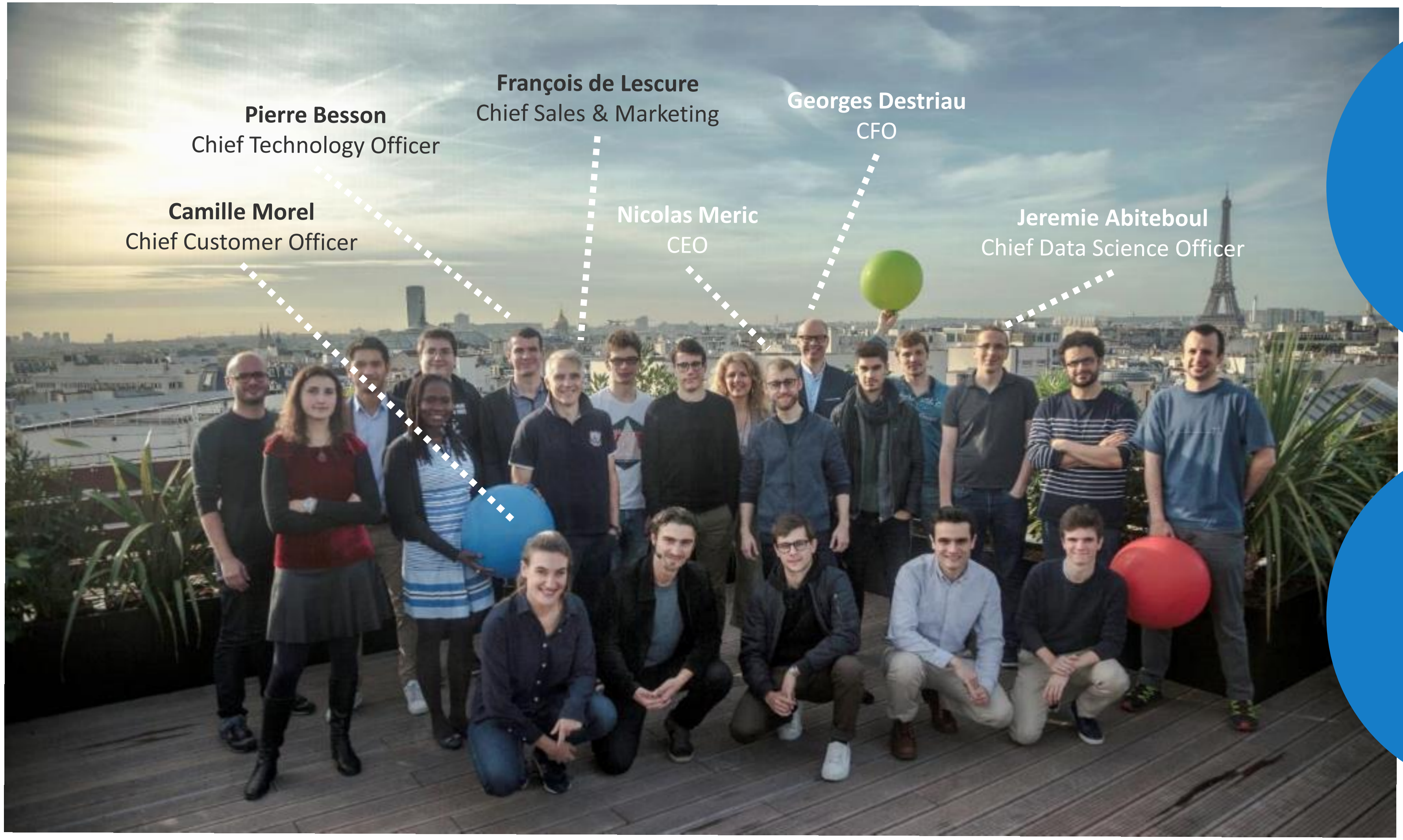
The FinTech50 2018 Hot Ten

PARIS - JUNE 2018 AI AWARDS IN FINANCIAL SERVICES Aurexia CONSULTING

RED HERRING WINNER 100 EUROPE

Finance Innovation FIN&TECH OF THE YEAR 2017

The success of a team

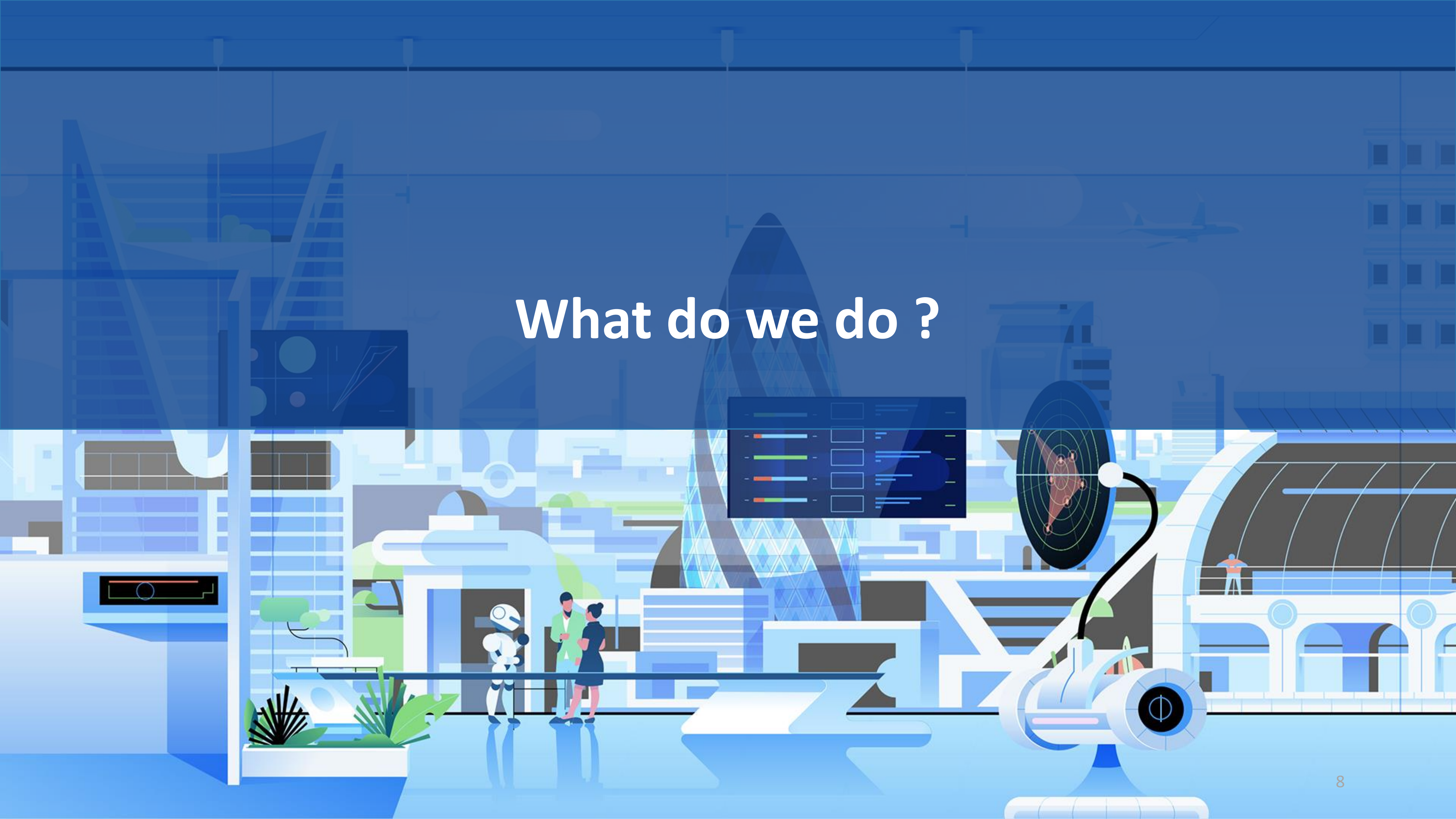


Now a team of

25

Series-B competed

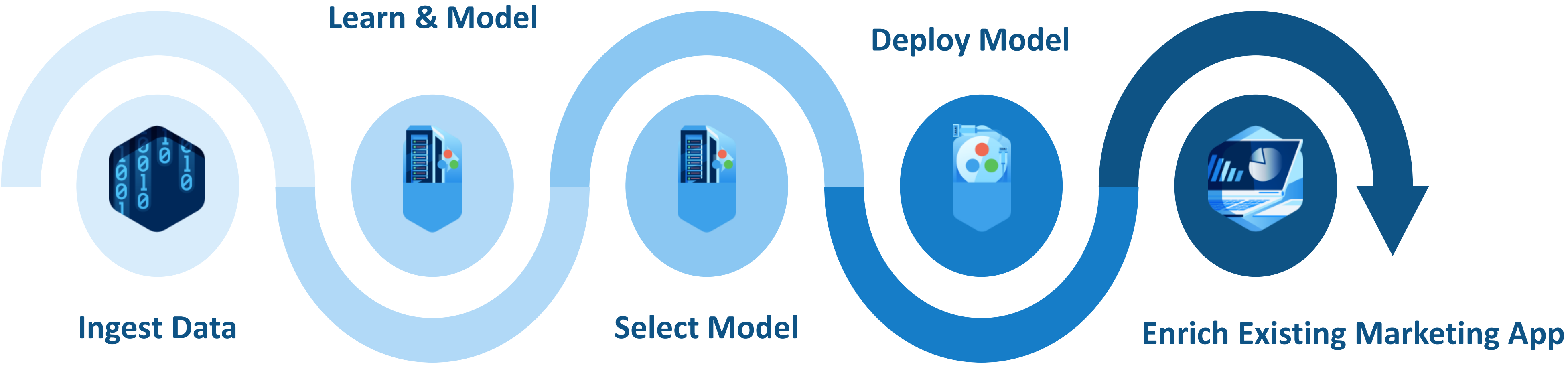
What do we do ?



Leverage data to reinvent the customer experience

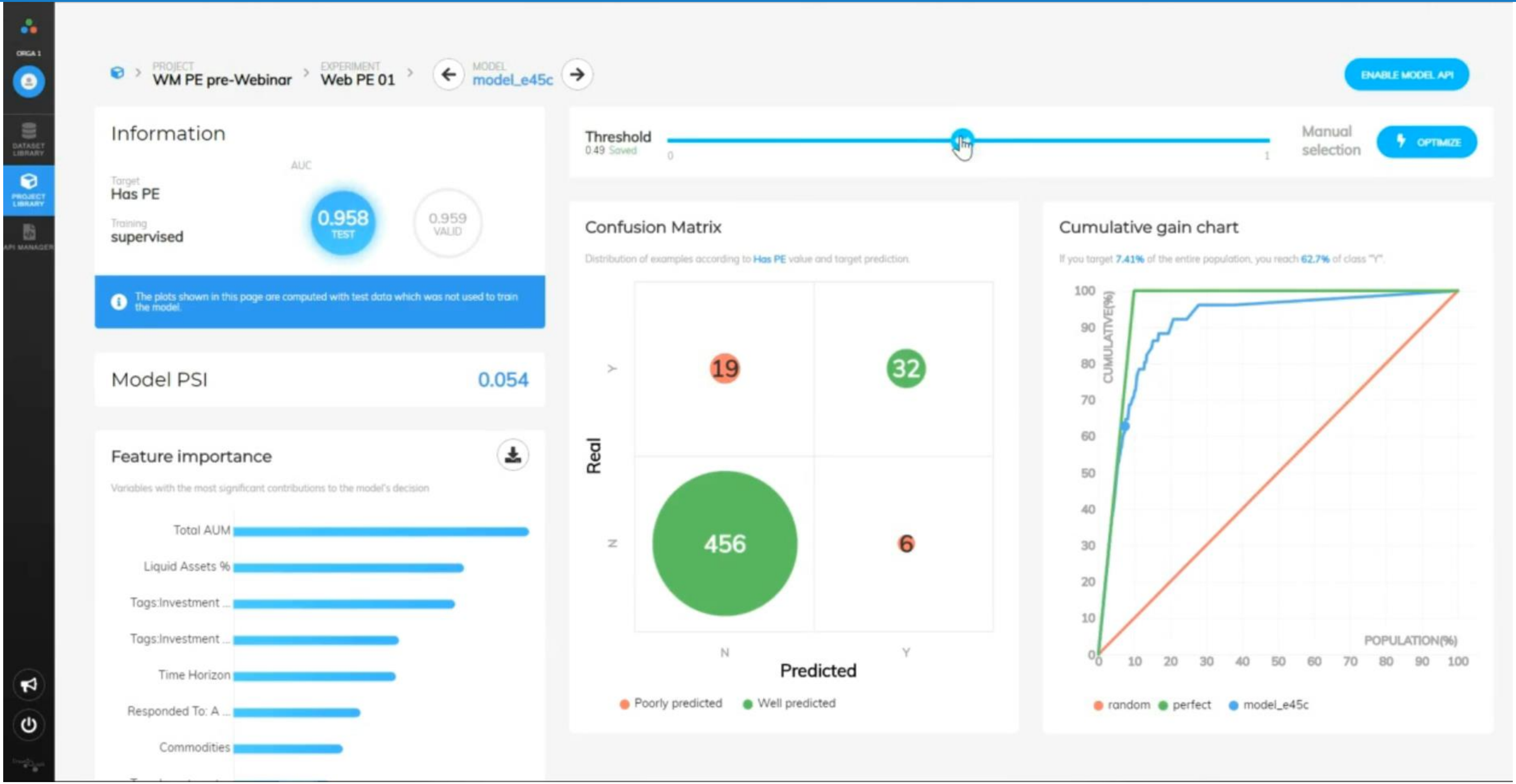
To make it a reality in large scale production, DreamQuark created an AI platform for business experts:

Simple & zero-code	Enterprise grade IT	Top data-driven decisions
On-prem or SaaS	Multi-tenant	Transparent & trustworthy



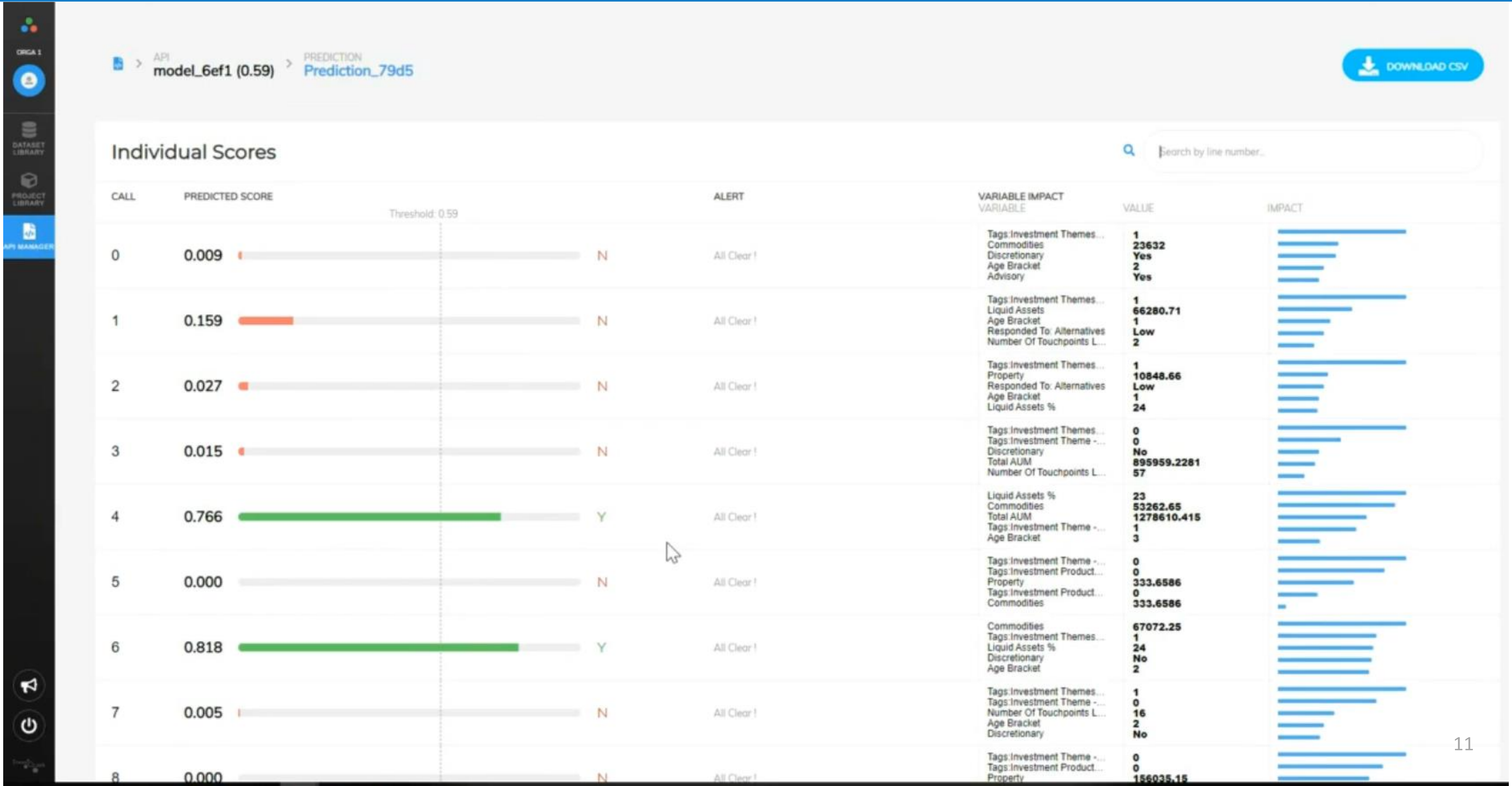
Empower business analysts with predictive analytics => Citizen Data Scientist

Easily generate predictive models

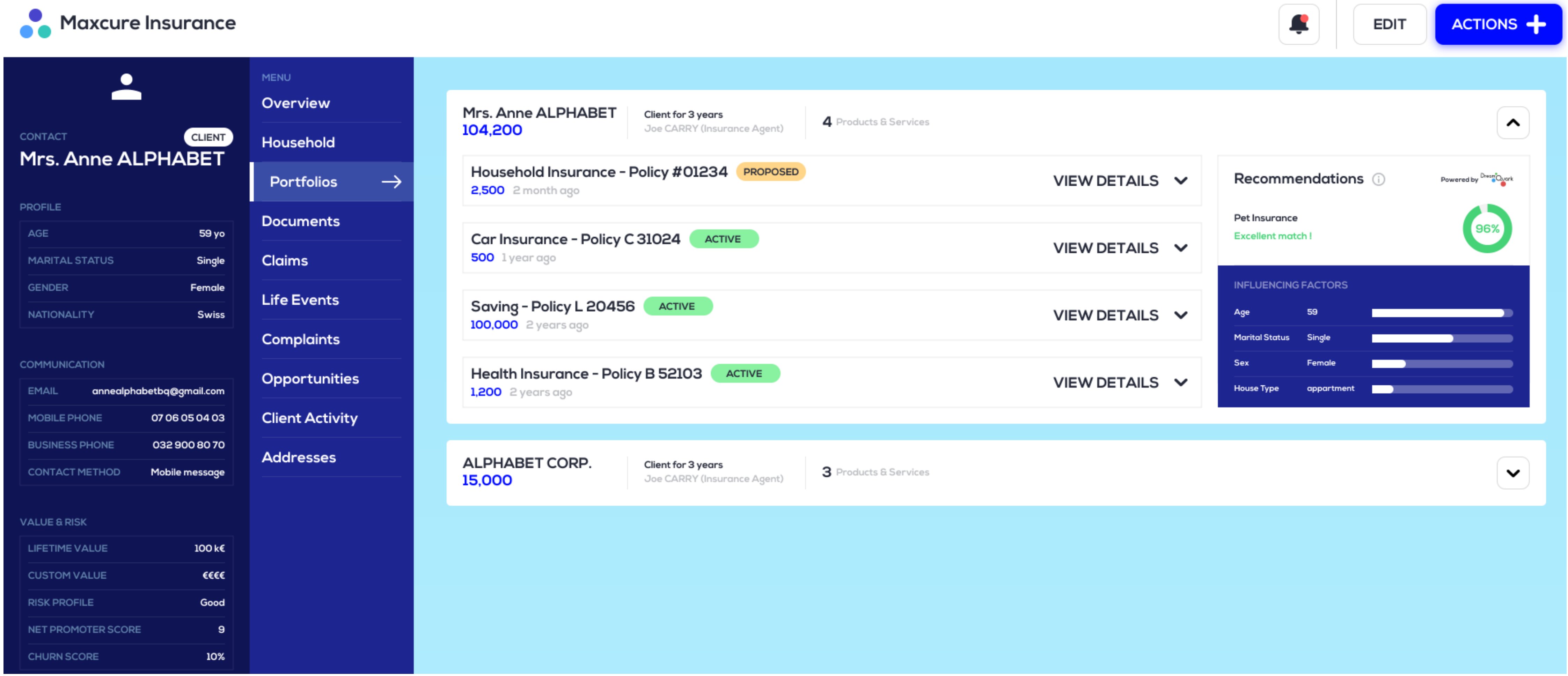


DreamQuark

Create actionable predictions



Integrate with existing business apps



Why is this interesting for (former) physicists ?

- Automatically training neural networks on tabular data
- Explainable AI: “opening the black-box”
- Fairness of AI systems
- Robustness of AI systems
- Challenges for AI systems in production

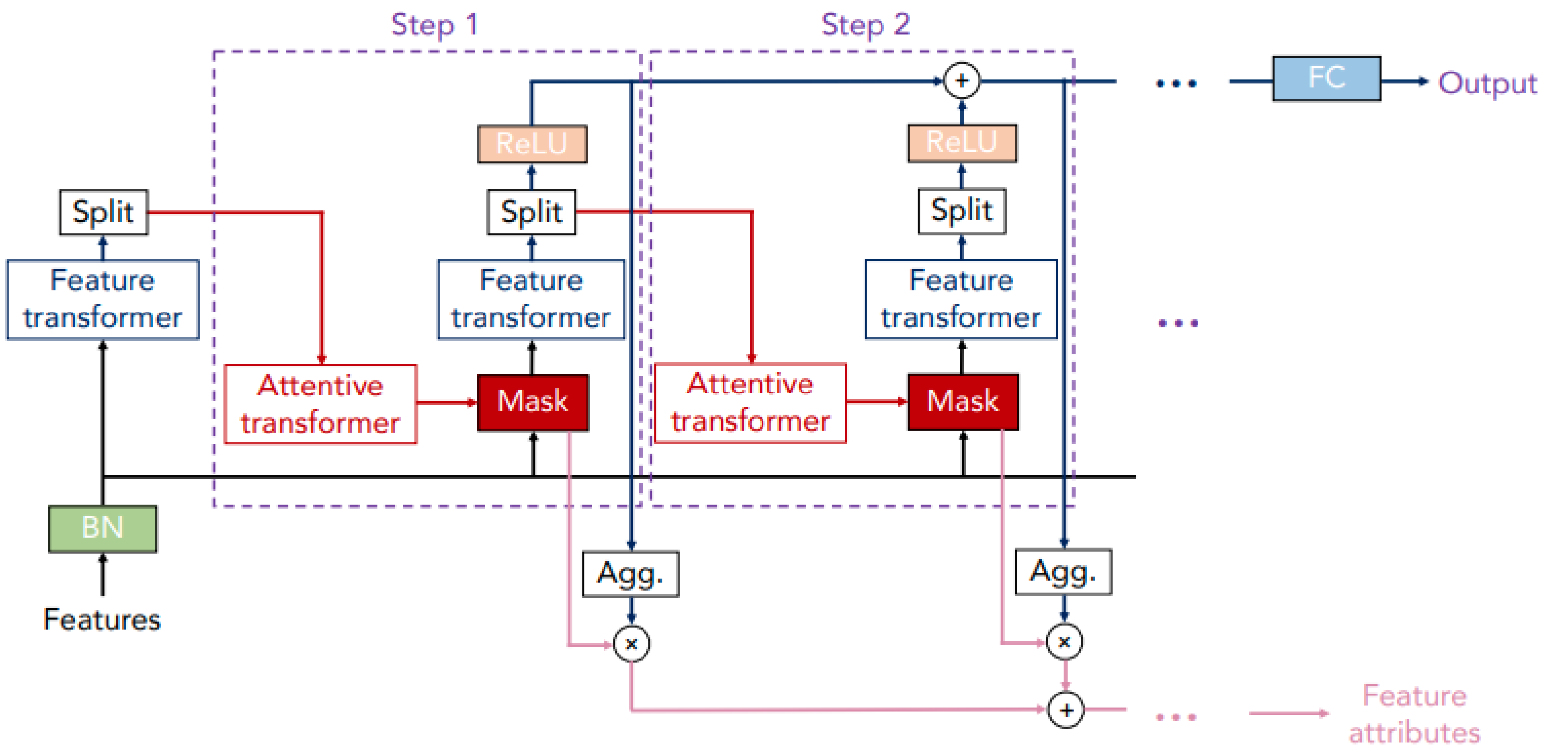
Training neural networks on tabular data

Neural networks for tabular data

- Most popular applications of neural networks focus on “unstructured” data (images, text, audio, video)
- Tabular data is actually less “structured” than e.g. images
- Most of the AI community still uses standard ML methods (boosting algorithms, random forests, etc) for tabular data
- We’ve found that neural networks are competitive if
  - Design suitable architectures (not necessarily very deep)
  - Find a reasonable grid for hyperparameter search [patent pending]
- Bonus: ensembling of neural networks and standard ML

TabNet: Attentive Interpretable Tabular Learning

[S.O. Arik & T. Pfister, arXiv September 2019]



- Sparse *learned* feature selection
- Sequential multi-step architecture (each step contributes to a portion of the decision)
- Attentive transformer ensures different steps learn from different features
- Non-linear processing of selected features (=> model capacity)

- Original publication by Google Cloud AI team: https://arxiv.org/abs/1908.07442
- Open-source (pytorch) implementation by DreamQuark: https://github.com/dreamquark-ai/tabnet

Explainable AI

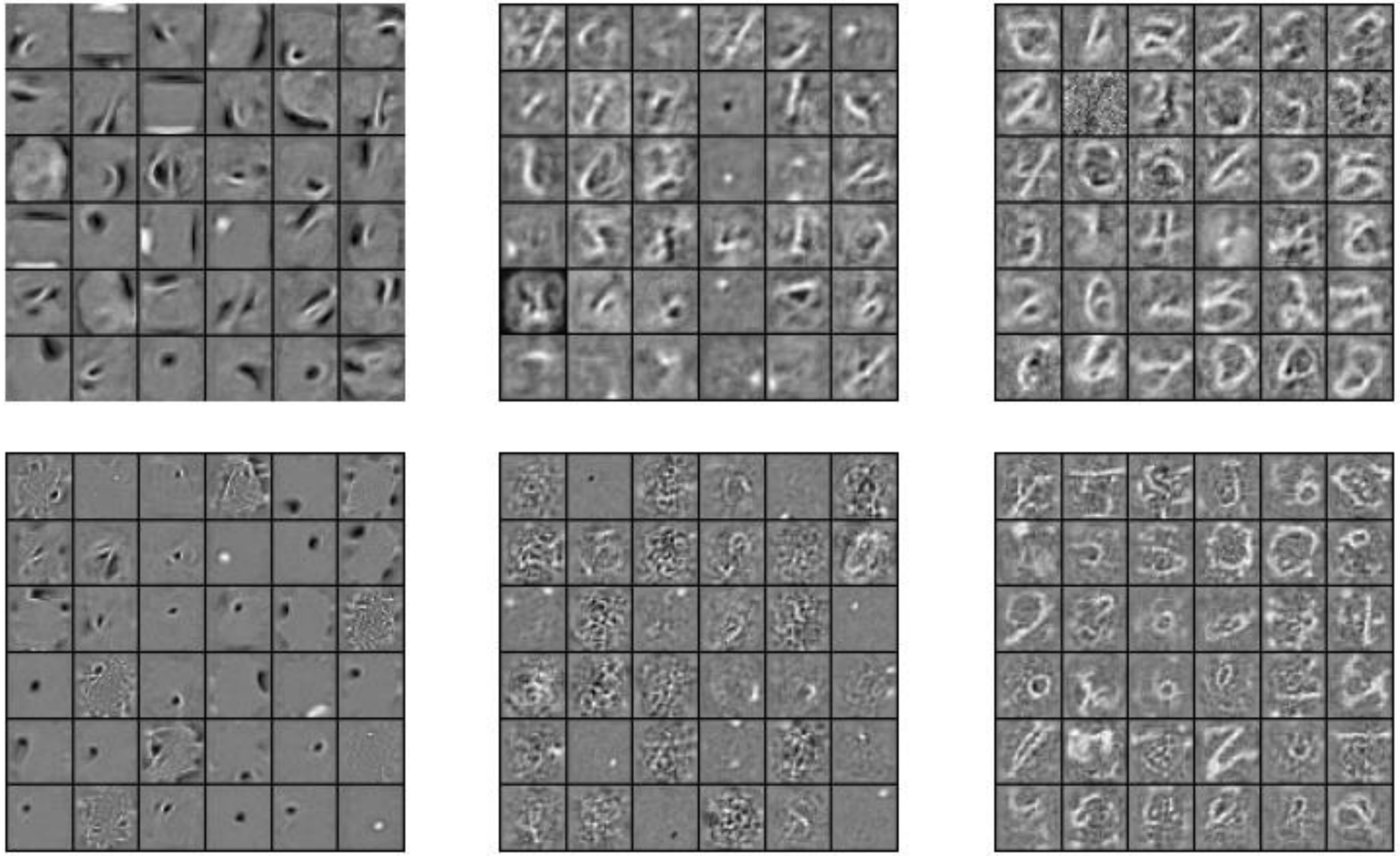
Explainable AI

- Understanding model decisions is a requirement for our users
  - Business users must have **confidence** in models before deploying them
  - **Regulations** require that decisions based on algorithms be made “understandable by non-experts”
  - For marketing use cases, explanations also make decisions **actionable**

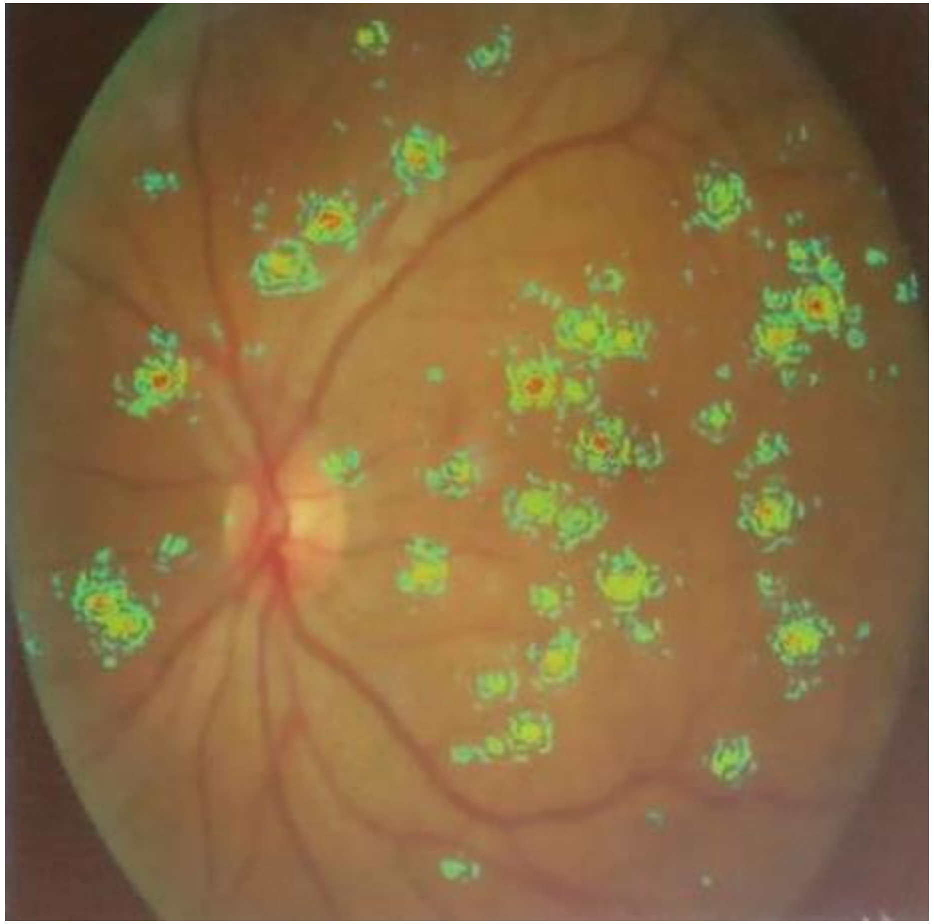
CALL	PREDICTED SCORE (Threshold: 0.32)		ALERT	VARIABLE IMPACT: VARIABLE	VALUE	IMPACT
0	0.639	Y	All Clear !	Tags:Investment Theme... Property Commodities Total AUM Tags:Investment Theme...	1 99193.84 123992.3 2730231.85 0	
1	0.648	Y	All Clear !	Property Tags:Investment Theme... Total AUM Fixed Income Tags:Investment Theme...	88167.32 1 2075721.9433 154292.81 1	

- Also much academic interest around this research topic
  - DARPA XAI program: https://www.darpa.mil/program/explainable-artificial-intelligence

Explaining Convolutional Neural Networks: saliency maps



Erhan, Bengio et al. 2013



DreamQuark 2016

Explaining Recurrent Neural Networks: cell activations

[Karpathy et al, 2015]

What about tabular data ?

- Several approaches stand out in the community:
  - **LIME** [Ribeiro et al, 2016 & https://github.com/marcotcr/lime]
    - ➢ Model agnostic, generates and scores random points around an example
    - ➢ Trains a simple (e.g. linear) model to reproduce this local behavior
  - **SHAP** [Strumbelj et al, 2014 & https://github.com/slundberg/shap]
    - ➢ Game theory approach: “connects optimal credit allocation with local explanations using the classic Shapley values from game theory”
  - **Permutation** Feature Importance

- DreamQuark method for individual explainability [patent pending]
  - For each ("local") prediction, use the neural network's underlying architecture to determine (locally) the most important variables for the model's decision
  - Decorrelate input variables in order to present more consistent explanations
- Definitely not the end of the story here, research is ongoing

Challenges for real-life AI systems

Fairness

- What constitutes fairness ?
  - *Demographic blindness* (training data is not correlated with demographics)
  - *Demographic parity* (outcomes are proportionally equal for all classes)
  - *Equal opportunity* (true-positive rates are equal for all classes)
  - *Equal odds* (true-positive and false-positive rates are equal for all classes)

- "impossibility theorems" of fairness, see e.g. [Zhao&Gordon, 2019]

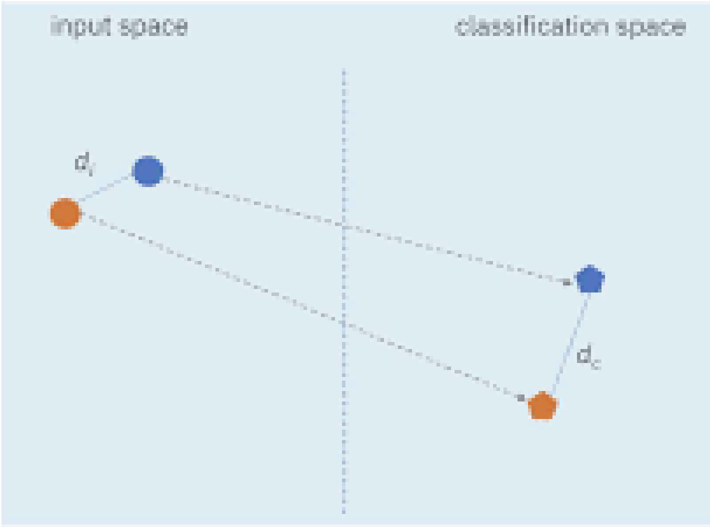
- *Equality of Opportunity in Supervised Learning* [Hardt et al, 2016]
  - Pre-define sensitive groups before training a model
  - => ensure no discrimination against protected groups in the learned model

Individual fairness = Robustness ?

- Definition of individual fairness [Dwork et al, 2011]

  => "similar individuals are treated similarly"
- In other words: a fair model is a robust model
- Difficulty: how do we define *similar* inputs ?

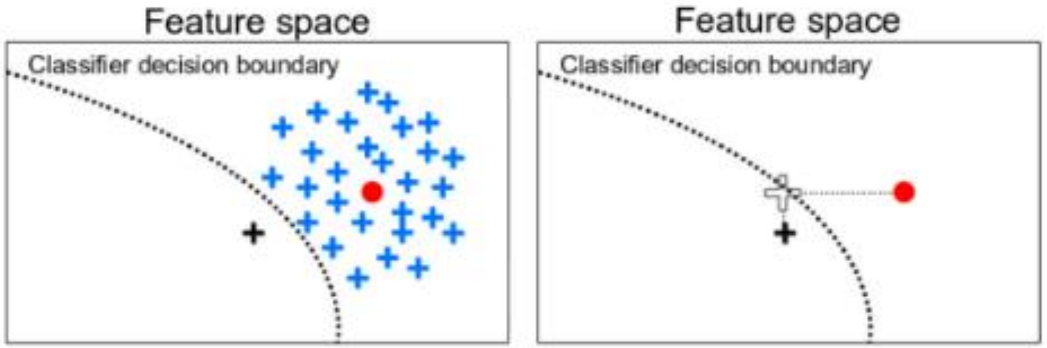
Fairness Through Awareness



Dwork et al. Fairness Through Awareness, 2011.

- Alternative approach: counterfactual explanations

  => Find the "closest" neighbor which the model predicts differently

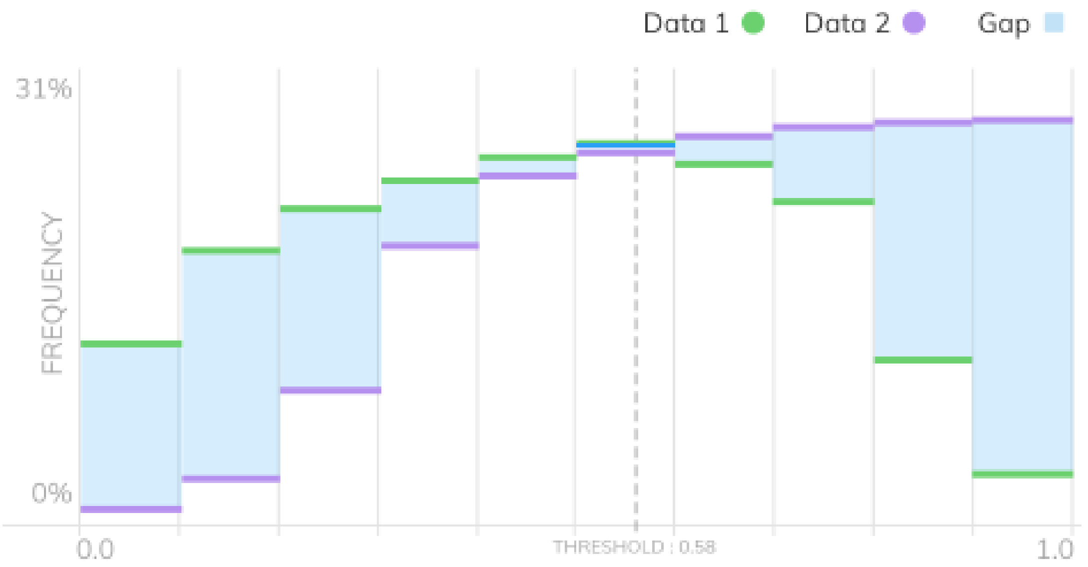


Step 1: Generation   Step 2: Feature Selection

[Lauger et al, 2017]
Method of « growing spheres »
for selecting sparse conterfactuals

Model stability in production

- Model performance (e.g. AUC, accuracy), even on *test data*, does not guarantee robustness and results in production
- => models bust me monitored in production, and eventually retrained
- **Model** stability (e.g. PSI)
- **Data** stability / data shifting
  - Does the distribution of datapoints in production match training data ?
  - Train a simple model to distinguish test data vs train data
  - If successful, your data has shifted



DreamQuark 2020

- Possible added benefit: robustness of AI system against adversarial attacks

Take-aways

Key take-aways

- Physicists can successfully transfer their skills and experience
- Innovation is possible in the industry, even as a small company
- As a private company: use open source software, develop proprietary code and file patents, but also publish open-source code

- Deep learning on tabular data is an emerging topic in AI community
- Beyond accuracy of models, challenges for both academia and industry
  - Explainable AI
  - Fairness
  - Robustness
  - Model stability and data stability

DreamQuark

Jeremie Abiteboul
Chief Data Science Officer
jeremie.Abiteboul@dreamquark.com

Appendix

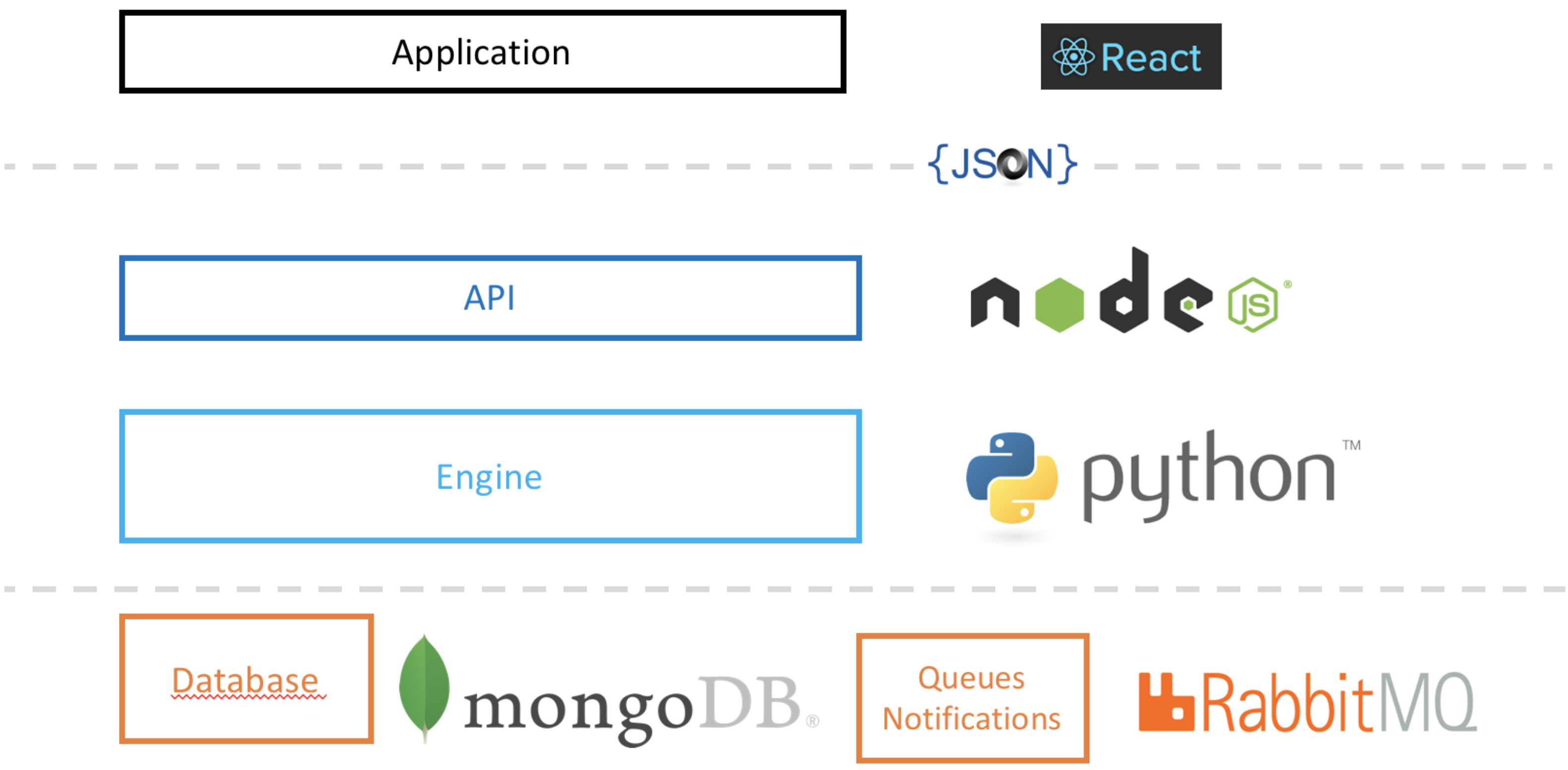
Automated generation of optimized tabular networks for structured data

[patent pending]

- Initial problem to solve:
  - Neural networks are a "universal approximator", but typically require significant manual effort for each new dataset
  - Many hyper-parameters: architecture, learning rate, regularization, etc

    => many-dimensional space for optimization, often carried out more or less manually
- Our solution to this problem, in a nutshell
  - Carry out training of a large number of models on a reduced dataset and/or for a small number of epochs, in order to gain "insights" on the dataset
  - => automatically deduce a reduced search space for hyperparameters
  - Optimize on this reduced search space using Bayesian methods

Technology stack



- + systematic use of Docker + Kubernetes for deployment
  (both cloud and on-premise)

DreamQuark

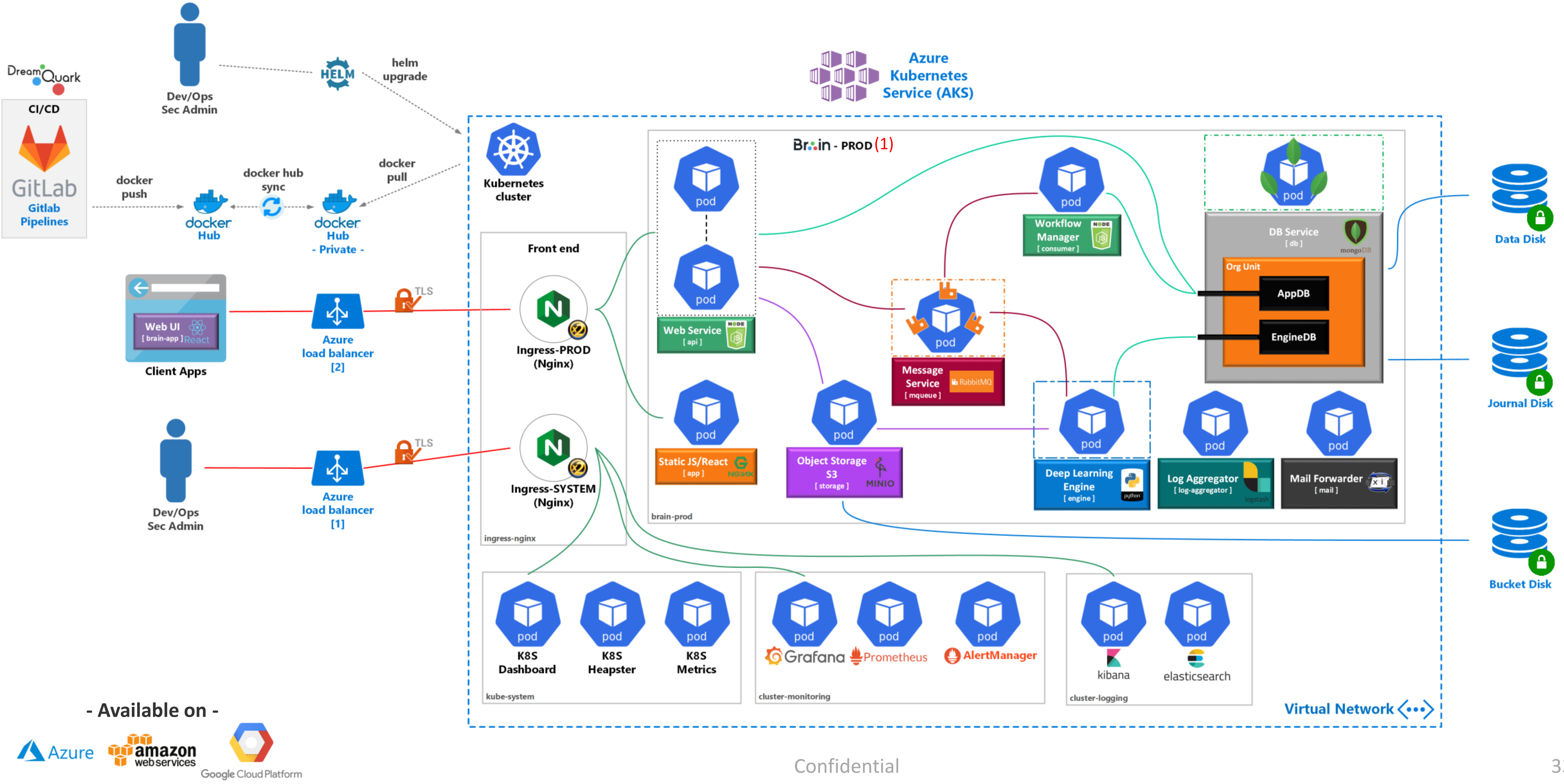
Containerization: docker + kubernetes

- Kubernetes: industry standard for container orchestration
- Deployment compatible with **multiple cloud providers & on-premise**

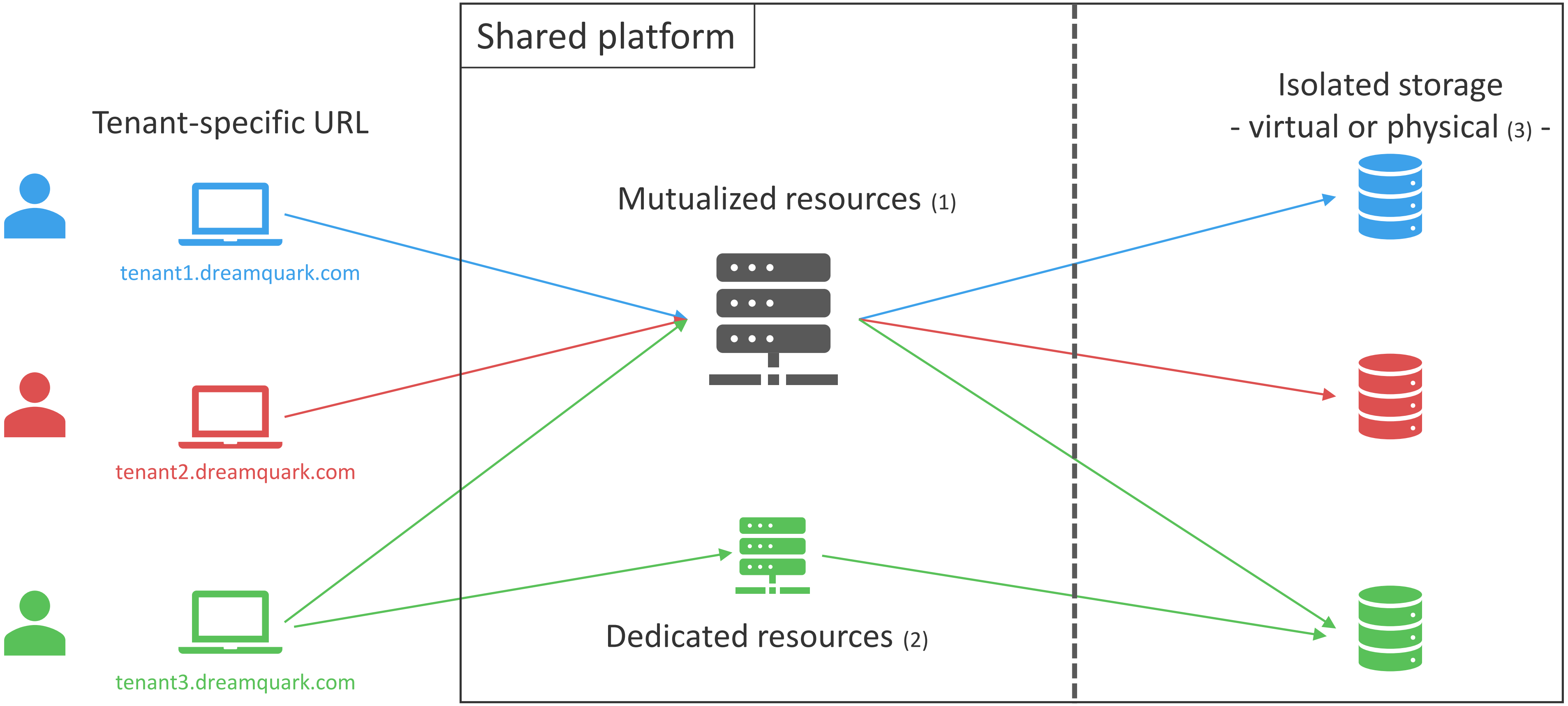
- Automated delivery of new releases
- Monitoring tools deployed as part of the product
- Easily manage many (independently scalable) micro-services

docker kubernetes

Architecture diagram



Multi-tenant platform



1. Instances of Brain micro-services (API or engine)
2. Only instances of engine micro-services
3. MongoDB: separated instances or separated databases

   S3 Object Storage: Separate Buckets or isolated "folders" in the same Bucket

Storage can be external and managed by the administrator