

**data
iku**

Alexis Fournier, AI Expert @Dataiku
alexis.fournier@dataiku.com

Agenda

Artificial Intelligence ?

Artificial Intelligence as an Enterprise Asset?

AI at Enterprise Scale: customer example

A small demo

Artificial Intelligence?

"the embryo of an electronic computer that [the Navy] expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence."

- NY Times 1958

Artificial Intelligence?

Advanced Analytics

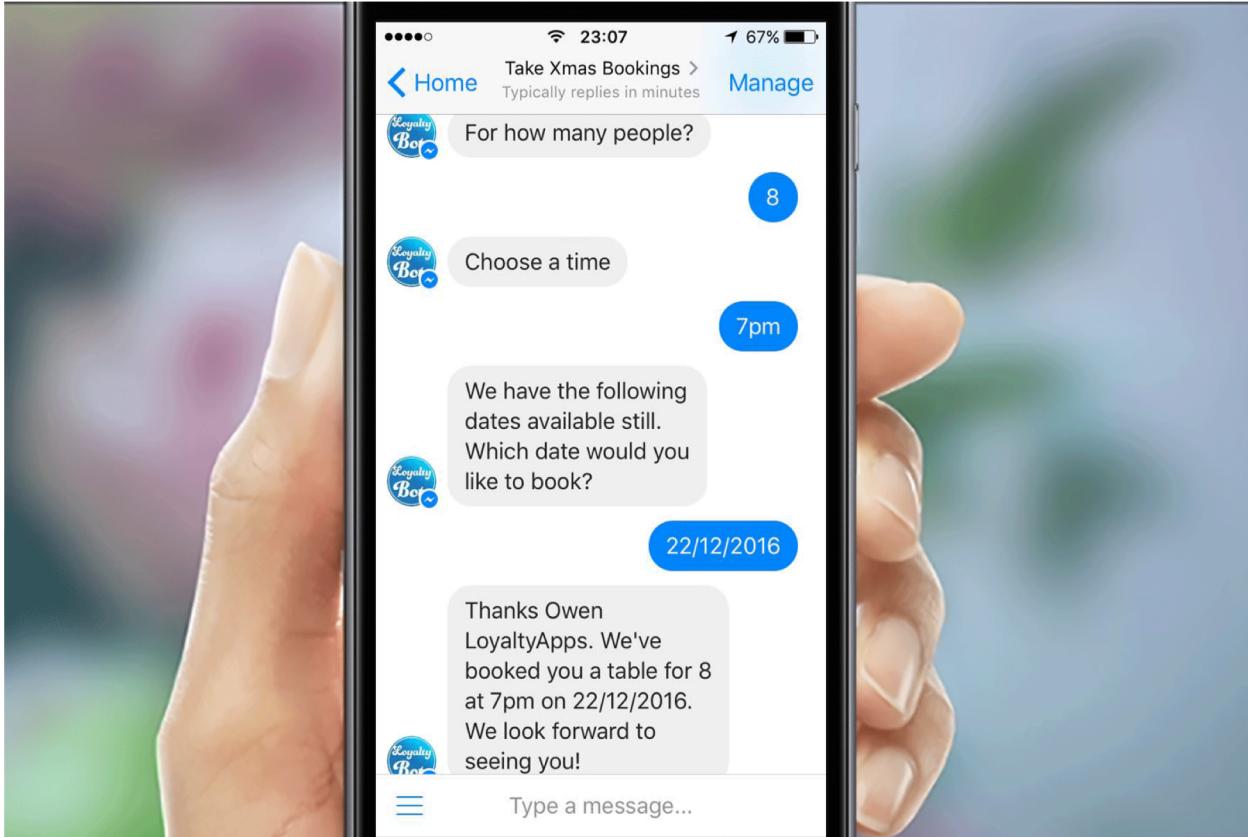
Data Science

Machine Learning

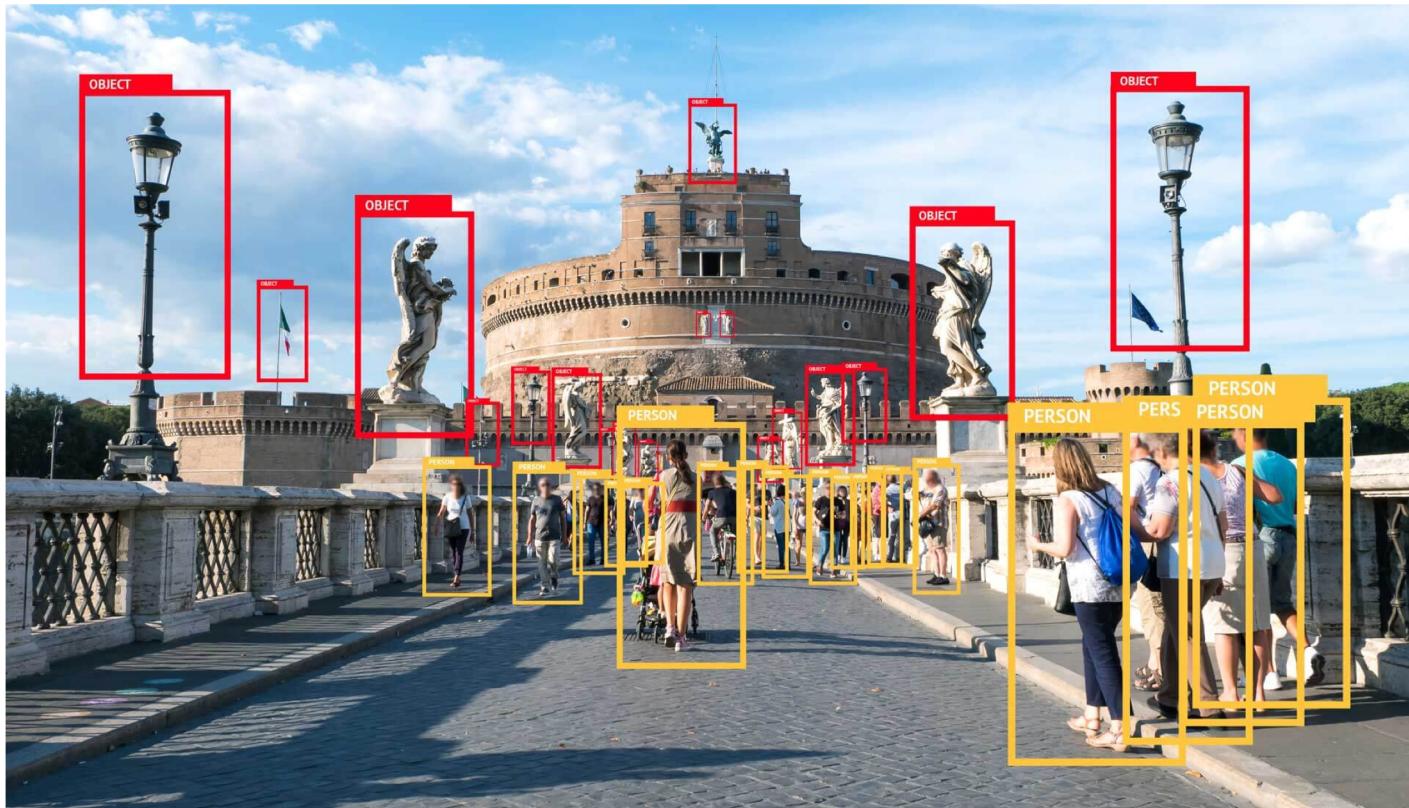
Big Data

Deep Learning

Artificial Intelligence?



Artificial Intelligence?



Artificial Intelligence?



And ... for Enterprise ?

Augmentation of process, product or decision based on the adequate processing of large amounts of data

Enterprise Artificial Intelligence

Why no later than today?

88% of companies obtain very little or no value from their data¹

companies that do experience

+30% EBITDA²

3X return on equity³

Artificial Intelligence as an Enterprise Asset?

Enterprise AI into an Enterprise Asset

What is the actual asset?

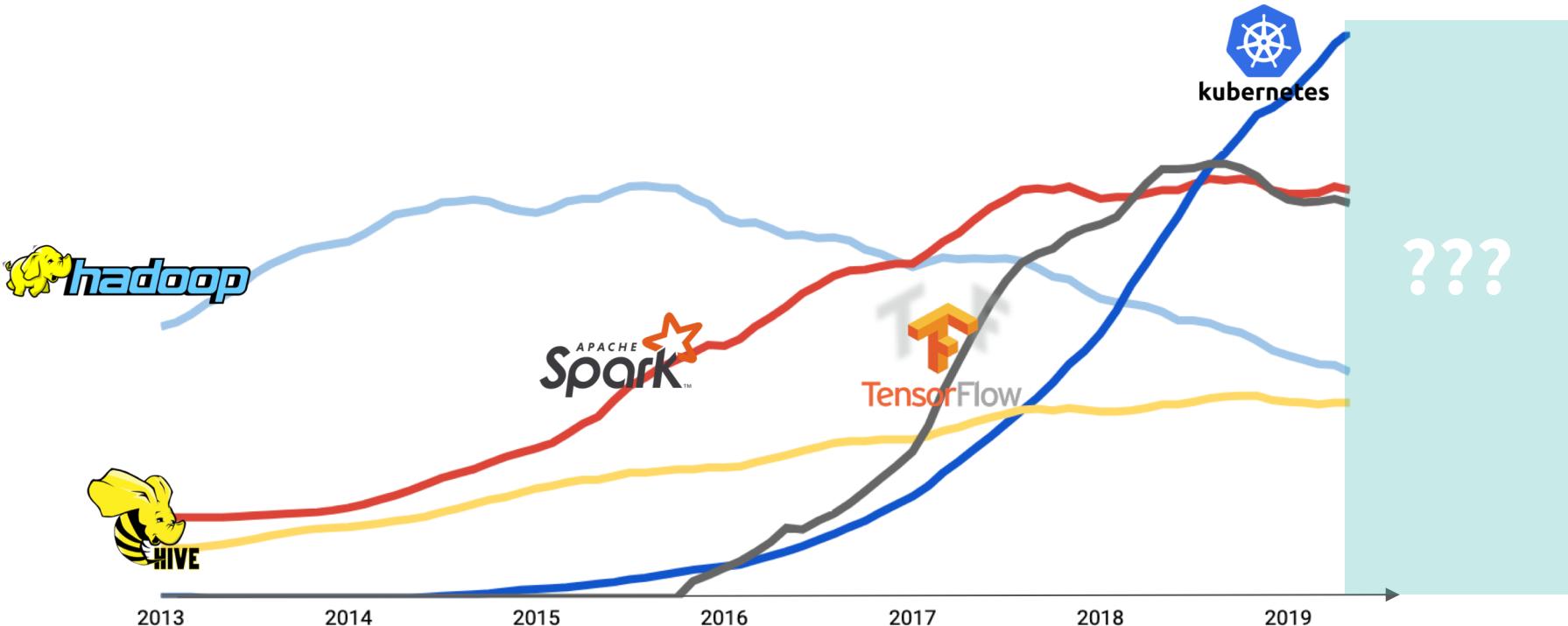
Enterprise AI into an Enterprise Asset

What is the actual asset?

Technology?

Enterprise AI into an Enterprise Asset

Technology as an enterprise asset?



Enterprise AI into an Enterprise Asset

What is the actual asset?

~~Technology?~~

Data?

Enterprise AI into an Enterprise Asset

Data as an Enterprise Asset

2.5 quintillion bytes of data created each day

(2.5 Millions T)

90% of data in the world generated in 2 years

Obsolescence of data, trends, formats...

Enterprise AI into an Enterprise Asset

What is the actual asset?

~~Technology?~~

~~Data?~~

Models?

Enterprise AI into an Enterprise Asset

What is the actual asset?

~~Technology?~~

~~Data?~~

~~Models?~~

Enterprise AI into an Enterprise Asset

What is the actual asset?

The most valuable asset is **the delivery of impactful models**

Data?

Models?

Enterprise AI into an Enterprise Asset

What stands in the way?

What is the greatest barrier(s) for companies when it comes to AI adoption?
(select up to three)



Operationalization

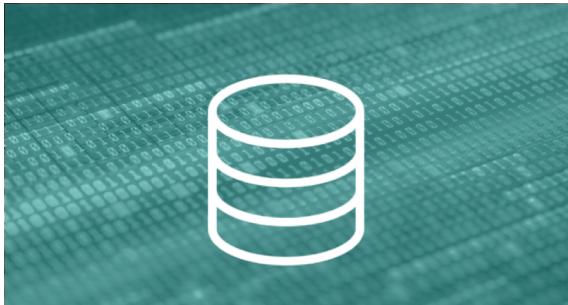
Governed AI integrated within operational systems

Inclusivity

Leveraging jointly AI/IT/Business talents

Enterprise AI into an Enterprise Asset

The Fundamental Challenge is not “Software”



Siloed Data/Tech



Siloed People



Siloed Processes



AI at Enterprise Scale: A customer example

AI At Enterprise Scale



1,800+

Users of the self-service
analytics system



130

Published datasets
in the last year



2000+

Data products created since
March 2017 (200+ in
production)



568

Automated Projects

Starting point: the experts

Where did it start?

The Digital League

Cross-functional team (supply chain, finance, IT, and engineering lines of business) to deliver data-driven products

Help break down silos

Emphasize collaboration

Drive infrastructure decisions

The roadblock: How to scale?

The idea: What if business owners could deliver their own
data innovation?

The team: Self-Service Data (SSD)

Scaling execution: Inclusive Data Science

Inclusive through education

The Digital Data Analyst (DDA) Curriculum

Training 1: Self-guided & videos

To get a basic understanding of tools and programs

Training 2: Intensive, week-long course.

Teaches digital data tools, DS, process excellence.

Over 1,000 graduates since January 2018

Training 3: A quarter-long program. In depth Training 2

Training for Executives: 1 day training on the value proposition of the SSD program.

Secure high level buy-in

Inclusive through tooling

The full stack available following DDA training

Dataiku: designing and deploying data products

Alation: data cataloging and search

Daashboard: in-house tool built for monitoring the state and status of all data products

Spotfire: visualization of final data product for end business users

Scaling execution: Operationalization

Moving from Raw Data to business output : **Automate** Production

1. **DATASETS ARE AVAILABLE** TO THOSE WITH APPROPRIATE ACCESS AND TAGGED BY TYPE FOR SIMPLE CATALOGING (raw data / transformed / published / consumption ...)
2. SELF-SERVICE DATA USER **CAN USE** ANY DATASET TO **CREATE PROJECTS IN A DESIGN** ENVIRONMENT
3. USER CAN **PROMOTE FROM THE DESIGN TO PRODUCTION** ENVIRONMENT AT THE CLICK OF A BUTTON
4. **AUTOMATIC CHECKS** ARE RUN ON PROJECT COMPLIANCE WITH PRODUCTION RULES (USED DATASETS, SIZE, NOMENCLATURES...), IF PASSED IT IS **PUSHED TO THE PRODUCTION** ENVIRONMENT AUTOMATICALLY

Scaling execution: Operationalization

Monitoring Execution : Automate Failure handling ?



1. Email automatically sent to Self-Service Data Team & data product owner



2. The system automatically tries to rerun the scenario(s)



3. If automatic rerun fails, the owner is notified automatically to fix the problem

AI at Enterprise Scale

Sample use cases of the 1,800+ community



Engineering

is using data from these tools to redesign parts and build jet engines more efficiently.



Supply chain

is using it to get better data insights into their shop floors and streamline supply chain processes.



Finance

is using it to understand key metrics such as cost, cash, etc.



The commercial group

(by leveraging data scientists) is using these tools to transform engine sensor data from customers and build analytics services from them.

Closing Tips

Tools do not replace training and continuing education

Gamification goes a long way

Enable users through failures with **proper error documentation**

Self service is **not a one time project** to be launched and forgotten

A central platform ensures virtuous reuse and accelerated value creation

Dataiku Home page

Dataiku DSS  Search items...

MY ITEMS See all > 

 Correlations analysis on revenue loss (dataiku) 
Predicting churn for Dataiku 

 compute k61niQj3 
Predicting churn for Dataiku 

 revenue loss 
Predicting churn for Dataiku 

 unlabeled customers churn prediction 
Predicting churn for Dataiku 

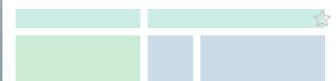
 Segments mod 
Predicting churn for Dataiku 

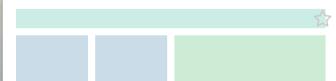
PROJECTS See all > 

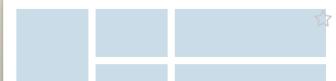
+ NEW PROJECT

 Predicting churn for Dataiku... 
Segmenting telecom customers, understanding churn, and scoring new customers.
 D +1
Sandbox 

DASHBOARDS See all > 

 Forecasting Customers that are likely to churn 
Predicting churn for Dataiku 1 page, 11 tiles 

 Segments Analysis 
Predicting churn for Dataiku 1 page, 8 tiles 

 Descriptive Statistics on the Customers 
Predicting churn for Dataiku 1 page, 10 tiles 

WIKIS See all >



Project overview

Predicting churn for Dataiku

Segmenting telecom customers, understanding churn, and scoring new customers.

D | A

sample_project sample

Sandbox

Flow Lab Dashboards Wiki Tasks

8 DATASETS 9 RECIPES 2 MODELS 4 NOTEBOOKS 3 ANALYSES 3 DASHBOARDS 1 ARTICLE 0 TASKS

GO TO FLOW

In this project you'll learn how to combine several models to build a churn prevention pipeline: segment customers, create insights to understand churn, and build a model to score new customers.

The data is from a major telecom operator. Just like pretty much any company in the world, they're concerned with keeping our customers happy, so they won't leave. In other words, they're looking at ways to reduce churn. To do this, they set up a task force of data analysts and people from business teams who came up with several business goals to reduce churn.

+ Add a todo list

WATCH 1 STAR 0

Correlations analysis on... 1 day ago

compute_k61niQj3 1 day ago

Forecasting Customers ... 1 day ago

revenue_loss 1 day ago

unlabeled_customers_... 1 day ago

TIMELINE

YESTERDAY

You edited recipe 09:24

compute_k61niQj3

You created jupyter notebook 09:23

Correlations analysis on revenue_loss (d...)

WEDNESDAY, 4 SEPTEMBER

You edited recipe 10:53

compute_k61niQj3

You created jupyter notebook 10:51

PCA on customers_within_segments (da...)

You edited metadata on project 10:46

Predicting churn for Dataiku

Todo list: Todo list

You edited tasks on project 10:46

Predicting churn for Dataiku

faire le model

Analytical flow

Predicting churn for Dataiku > Flow

Search All 9 recipes 1 folder 2 models 8 datasets

+ RECIPE + DATASET

compute_k61niQj3

DETAILS DISCUSSIONS ACTIONS

Edit Run Tag Unwatch Star Delete

The analytical flow diagram illustrates the data pipeline for predicting customer churn. It starts with two main data sources: 'customers' and 'unlabeled_customers'. The 'customers' source feeds into a 'customers_prepared' step, which then splits into two parallel paths. One path leads to a 'Segments model (KMEANS)' step, resulting in 'customers_within_segments'. The other path leads to a 'Churn prediction model (RANDOM_FOREST)' step, resulting in 'unlabeled_customers_churn_prediction'. Both segments and churn predictions are aggregated into a 'revenue_loss' step. Finally, the 'revenue_loss' step feeds into a 'PowerBI' report.

```
graph LR; customers[customers] --> customers_prepared[customers_prepared]; customers_prepared --> segments[Segments model (KMEANS)]; customers_prepared --> churn[Churn prediction model (RANDOM FOREST)]; segments --> customers_within_segments[customers_within_segments]; churn --> unlabeled_churn[unlabeled_customers_churn_prediction]; customers_within_segments --> unlabeled_churn; unlabeled_churn --> revenue_loss[revenue_loss]; revenue_loss --> powerbi[PowerBI]
```

View · default FLOW ACTIONS

Data Connexion

Predicting churn for Dataiku                          

New dataset

 Files

Upload your files
Server Filesystem
Files in folder

 Hadoop

Hadoop connection is not enabled on your DSS instance. Please contact your administrator

 SQL

MySQL	Greenplum
PostgreSQL	Teradata
Vertica	Oracle
Amazon Redshift	MS SQL Server
Snowflake	SAP HANA
IBM Netezza	Google BigQuery
Other SQL	

 Cloud Storages

Amazon S3	FTP
Azure Blob Storage	SFTP
Google Cloud Storage	SCP
	HTTP
	HTTP (with cache)

 NoSQL

MongoDB
Cassandra
ElasticSearch

 Social

Twitter

 DSS

Managed dataset
Folder
Dataset from another project
Metrics
Internal stats
Editable

 Import existing

Choose connection to import from
Import from catalog

More dataset options?

[Browse plugins](#)

Export to Power BI

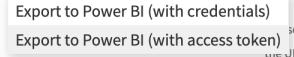
Predicting churn for Dataiku         Recipes

compute_k61niQj3 

Summary Settings Input / Output Advanced History        SAVED! ACTIONS ▾

Other exports

Exporter Export to Power BI (with access token) 

DSS project key  Use type the current DSS Project key (as seen in the URL above)

Dataset Alexis-Nantes-Bls  Power BI output dataset name

Overwrite dataset Whether Power BI dataset will be overwritten or not (or created in the case of a new export)

Table dss-data  Power BI output table name (will always be 'dss-data')

Buffer size 100  Number of records to send to Power BI at each write

Show all settings

Clear before export Clear the target partition before starting the export

Thank you!

(questions?)