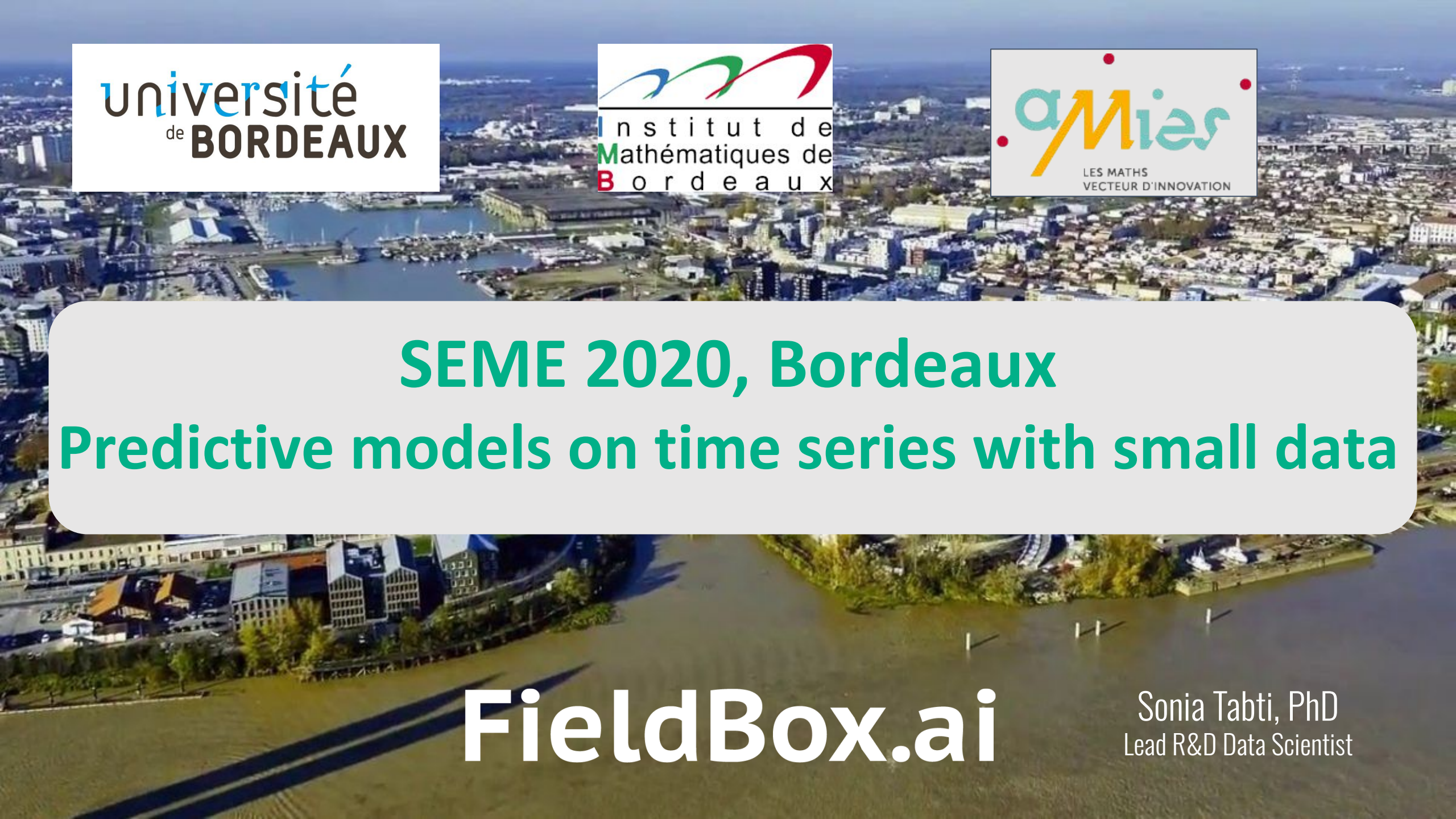


SEME 2020, Bordeaux

Predictive models on time series with small data

FieldBox.ai

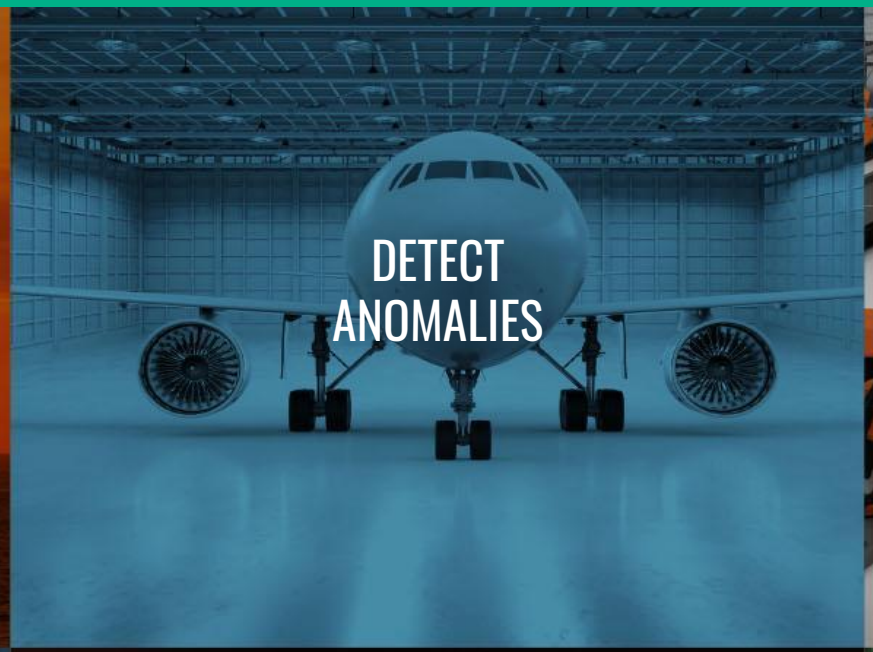
Sonia Tabti, PhD
Lead R&D Data Scientist

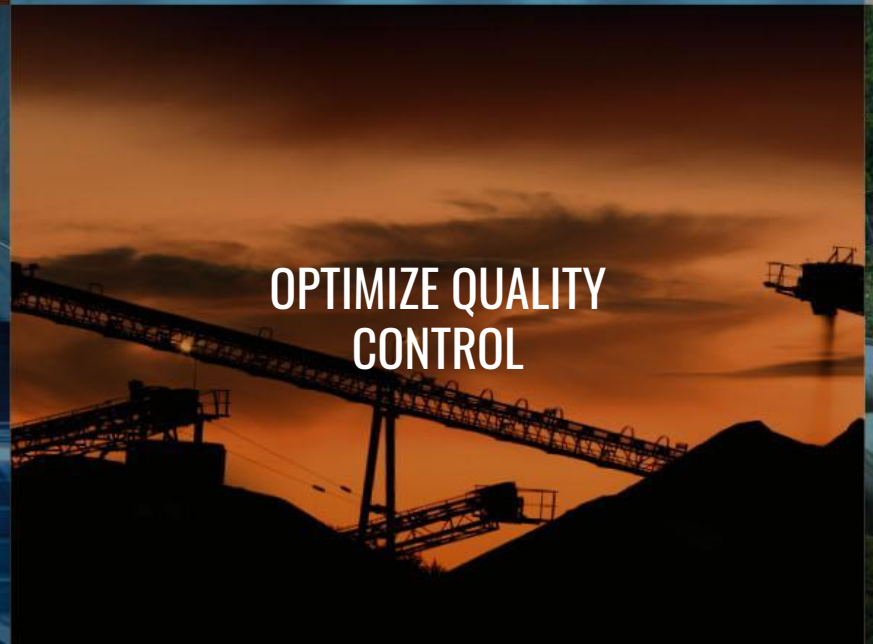
FieldBox.ai

Artificial Intelligence for Industry

Who are we?

USING ARTIFICIAL INTELLIGENCE TO REACH OPERATIONAL EXCELLENCE IN INDUSTRIAL OPERATIONS



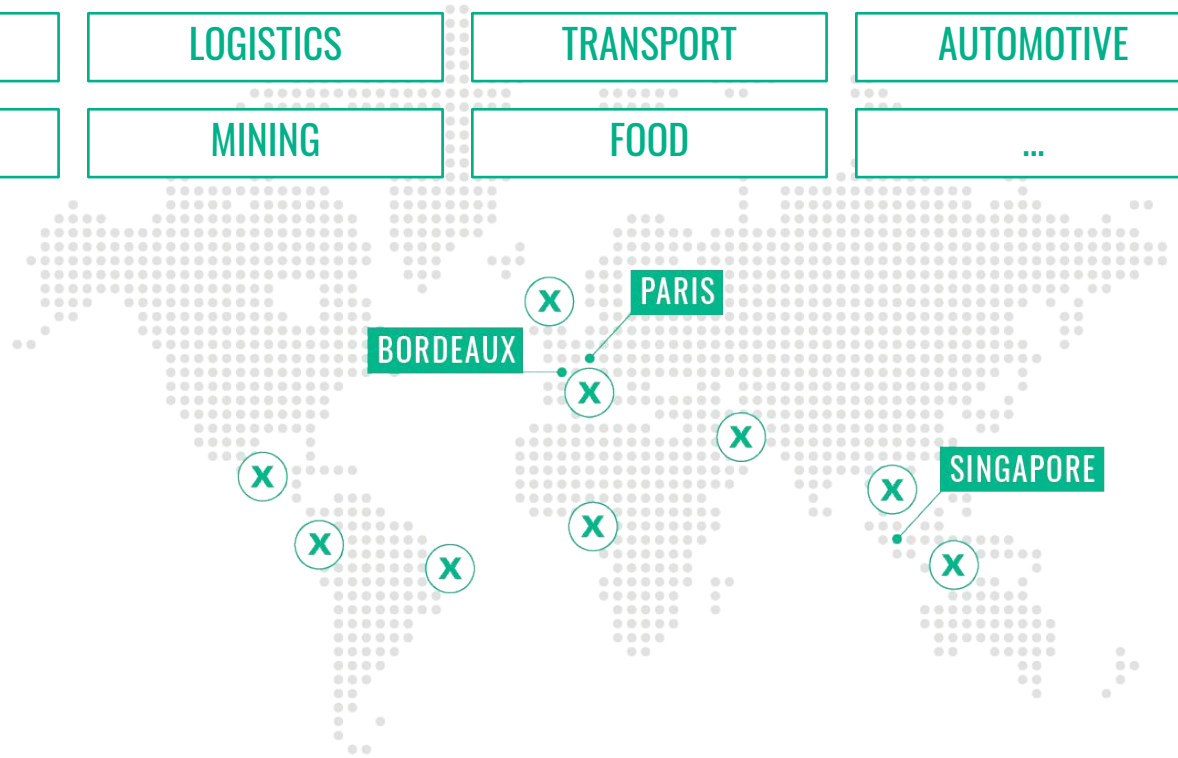


FieldBox.ai

CUSTOMERS AND OPERATIONS ON FIVE CONTINENTS SINCE 2014

ENERGY | LOGISTICS | TRANSPORT | AUTOMOTIVE | UTILITIES | MANUFACTURING

LUXURY GOODS | MINING | FOOD | ...



OUR OFFICES

CLIENTS' SITES OPERATED WITH FIELDBOX PLATFORM

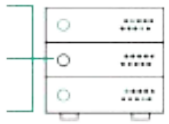
SELECTED CUSTOMERS

TOTAL | SBM OFFSHORE | SAINT-GOBAIN | vallourec | GROUPE ADP | SNCF | RICHEMONT | SUEZ | TEREGA

FieldBox.ai

GET THE MOST VALUE OUT OF YOUR INDUSTRIAL DATA WITH FIELDBOX SOFTWARE PLATFORM AND SERVICES

CONNECT
Manage Industrial Data

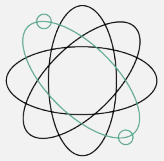


Collect data
from all industrial sources

Identify and qualify
legacy data sources

Develop ingestion analytics
to create dynamic data pipeline

ANALYZE
Leverage Data Science



Create algorithm
and train models

Data science and AI
model development

Data training for
Engineers and SMEs

DEPLOY
Capture ROI in Operations

Run AI Agent
autonomously & securely

Deployment and
Industrialization at scale

Predictive models
maintenance and support

SOFTWARE

SERVICE

FieldBox.ai

ONE TEAM

SOFTWARE ENGINEERING

~20 Engineers

Backend developers

Frontend developers

ML Engineering

DATA SCIENCE

~40 Data Scientists

M2
Engineer
PhD

Statistics
Machine Learning
Optimization
Physical modeling

Projects for clients
Product features
R&D

Time Series
Image processing
Natural language

Linear Regression
Random Forests
Convolutional Neural Networks
Variational Autoencoders

Small data
Big data

PROFESSIONAL SERVICES

~20 Engineers

Industry experts

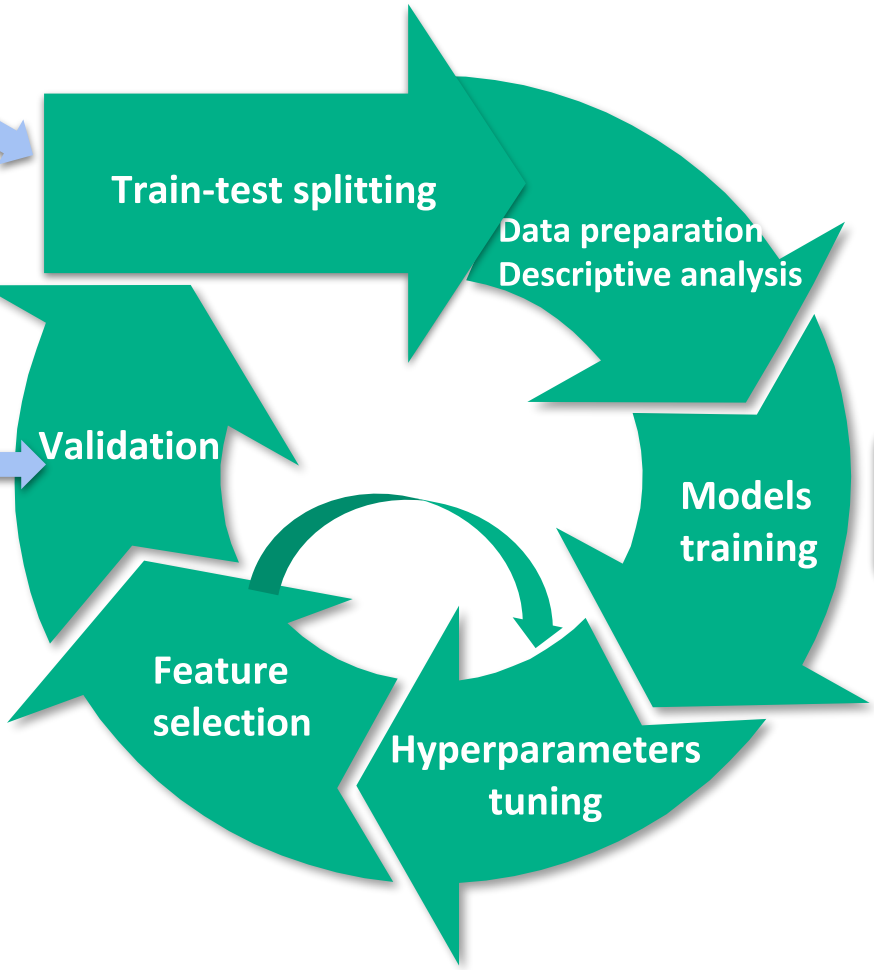
Project managers

Consultants

FieldBox.ai

"DATA SCIENCE" PROJECT LIFE CYCLE

-Formulate analytical question
-Collect data and metier expertise
-Visualize raw data



-Do we answer the question asked by the client?
-If not, reformulate or collect more data
-Then repeat the cycle
- If we do, build a nice interface to visualize results and deploy

-Build a baseline model first!
-Feature engineering

What is a typical use case of time series prediction ?

date	X_1	X_2	X_3	X_4	y
2018-11-13	192.23	46725710	191.63	197.18	191.4501
2018-11-12	194.17	50991030	199.00	199.85	193.7900
2018-11-09	204.47	34317760	205.55	206.01	202.2500
2018-11-08	208.49	25289270	209.98	210.12	206.7500
2018-11-07	209.95	33291640	205.97	210.06	204.1300
2018-11-06	203.77	31774720	201.92	204.72	201.6900
2018-11-05	201.59	66072170	204.30	204.39	198.1700
2018-11-02	207.48	91046560	209.55	213.65	205.4300
2018-11-01	222.22	52954070	219.05	222.36	216.8100
2018-10-31	218.86	38016810	216.88	220.45	216.6200

Suppose you have a clean historic dataset composed of features (X_i) and a target y

You want to predict the future values of y

What would be a baseline solution?

- You have many types of models to deal with this problem depending on the use case
  - Regressive and auto-regressive models (linear regression, ARIMA, …)
  - Tree-based models (Random forests regressor, Gradient Boosted trees regression, ...)
  - Neural networks models (1D-CNNs), LSTMs (Long Short Time Memory networks),...

- You have many tools in Python to build Machine learning models efficiently
  - Pandas to deal with dataframes
  - Scikit learn to build standard machine learning models
  - Keras/Tensorflow for neural networks models, PyTorch, ...

What if the dataset is too small ? …

- Big data VS small data

- When the dataset is small, you will quickly notice that:
  - Performances on the test set are not great …
  - Models overfit easily
  - “Simple” models can be better
  - Feature selection is crucial, …

What preliminary ideas do we have to tackle this issue?

- Data augmentation!
- Example of tool to explore: tsaug package

What do I expect from you?

- Quick survey on papers, best practices and existing tools in Python to deal with small data in this context
- Use a regular dataset that is not small, build some predictive models with it
- Simulate that it is small (maybe with various degrees), get new results with the models
- Explore and compare the results of some solutions you might find
  - You can start with data augmentation with tsaug for instance
  - Compare with the results obtained on regular dataset and small dataset
  - Extend to other datasets
- Write documentation and provide clean and commented code

Thank you for your attention!

Questions?


suez
REDUCE TIME FOR PIPE INSPECTION BY 50%
REDUCE REPORTING TIME BY 80%
Operations focused Cloud-to-Edge analytics stack
Understanding and collaboration with Subject Matter Experts
Integration in complex IT ecosystem
FieldBox.ai


PARIS AÉROPORT
LOST BAGS RATE
REDUCED BY 30%
MAINTENANCE
WORK ORDER
REDUCED BY 20%
Agent training on key historical data
Processing 2 issues concurrently
8 weeks to provide first ROI
FieldBox.ai