CONCRETE USE-CASES WITH REAL-WORLD DATA FOR ENERGY TRANSITION

G. Caputo, A. Mutschler | Deepki, Paris, France | www.deepki.com | giulia.caputo@deepki.com | aurelie.mutschler@deepki.com

Deepki

Deepki Ready is a SaaS that collects data and provides various tools that allow companies to have a better insight of their real estate and increase their energy efficiency. Deepki Ready relies on a wide range of functionalities : visualizations for exploration and comparison of sites, bill control, prices optimization, real-time alerts, assessment of gains, recommendation of actions ...

DATA @DEEPKI

FROM RAW DATA TO VALUABLE INFORMATION

1 - DATA COLLECTION

2 - DATA INTEGRATION

3 - DATA ANALYSIS

MULTIPLE SOURCES → RAW DATA → mongoDB (noSQL) / influxdb (Time-series) → Deepki Ready

COLLECT AND ANALYZE HETEROGENEOUS DATA

< PATRIMONIAL DATA >
Sites, addresses, surfaces buildings' description...

< ENERGY >
Load curves and gas/water consumption (from annual to fine-grained) prices, overloads

< ACTIVITY >
number of occupants/visitors revenue, equipments...

< OPEN DATA >
weather, opening hours, geocoding, taxes,...

data.gouv.fr

FROM MULTIPLE SOURCES

< IOT > API sFTP — LoRa, QUALISTED

< AMR > API sFTP — m2ocity, eveler

< CLIENT DATA > Forms XLS — XLSX

< PROVIDERS > web scrapping, pdf parsing — eau de Paris, direct energie, eni, EDF, CPCU

< OPEN DATA > API — World Meteorological Organization, Google Developers

< BMS / CTMS > API sFTP — SIEMENS, PcVue Solutions, EIFFAGE ENERGIE

< ERP / SOFTWARE > API sFTP — SAP, ORACLE JD EDWARDS ENTERPRISEONE

WORKING WITH REAL-WORLD DATA : CHALLENGES

QUALITY ISSUES : handling errors and missing values

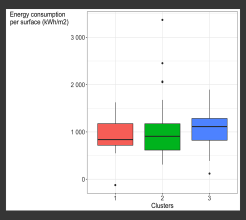
NOMENCLATURE & STRUCTURE : unifying databases while preserving each domains/client specificities

DEALING WITH DIFFFERENT TIMESCALES : static vs. dynamic data, handling updates, optimizing performances for time series

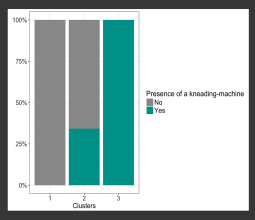
USE CASES

EXPLORE A REAL ESTATE AND PRIORITIZE ACTIONS FOR A BAKERY CHAIN

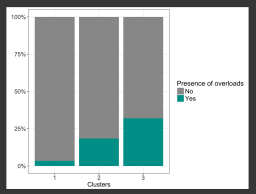
Understand the issues of a bakery chain, qualify what is a «usual» or « anomalous » behaviour as regards energy consumption, and explain what causes the anomalies.



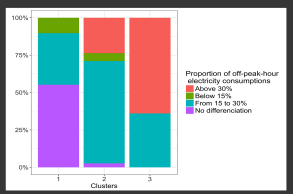
The median of energy consumption per surface was significantly higher in the third cluster



All bakeries from the third cluster were equiped with a kneading machine



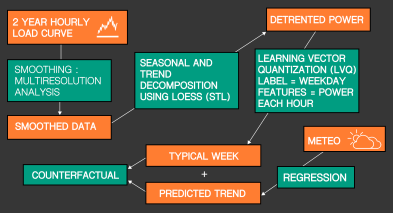
Bakeries from the third cluster were more likely to have overloads

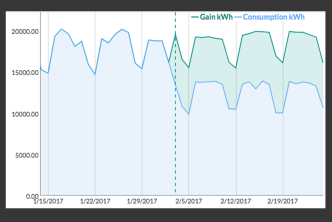


Bakeries from the third cluster consumed more during off-peak hours compared to the other clusters. We have evidenced that their ovens were maladjusted.

COMPUTE GAINS ASSOCIATED WITH A CHANGE OF EQUIPMENT IN A FACTORY

Determine the counterfactual after changing equipments in order to estimate gains.

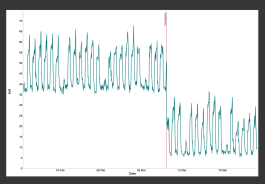




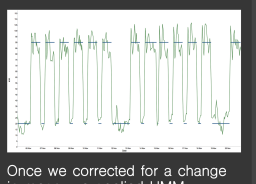
Gains estimate computed as the difference between actual power and its counterfactual in case no equipment was changed

ANOMALY DETECTION : REAL-TIME ALERTS IN INTERMARCHÉ STORES

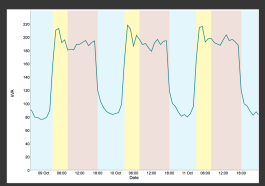
Plug a real-time anomaly detection system to detect electricity overconsumptions due to regulation or equipments issues in an Intermarché store.



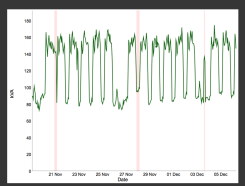
PELT algortihm was used to detect changepoint in time series



Once we corrected for a change in mean, we applied HMM algorithm considering 2 hidden states respectively open vs close.



We were able to distinguish several phases in the time series by crossing hmm results with opening hours



It enables us to detect different kinds of alerts, depending on which phase the problem occured. Here are several examples of over-consumptions during inactivity phases.

IN THE FUTURE

CONTINUOUSLY IMPROVE DATA QUALITY
- Missing data imputation
- Feature engineering : extract the most valuable information from raw data

INCREASE THE ROBUSTNESS OF THE METHODS USED
- Performance analysis of unsupervised methods
- Generalized metrics/processes for continuous monitoring of models' performances

TAKE ADVANTAGE OF ALL DATA AND MAKE THE SOLUTIONS MORE AND MORE UNIVERSAL
- DDK : a cross-clients, cross-domains database relying on a unified nomenclature and normalized data
- Inductive/transductive transfer learning : multi-task models, domain adaptation

TACKLE PERFORMANCE ISSUES
- online methods, cluster computing

CONSULTING
- client-specific methods/analyses
- strong reliance on expert knowledge
- use domain-specific kpis

BIG DATA
- universal models
- reproducible research
- make the most of all data types combined
- knowledge discovery