bip. xTech

Data Mining and Text Mining Course Project

November 2020

HERE TO DARE

BIP Overview

BIP in a “nutshell”

12 Countries

+3300 Professionals

+4500 Successful projects

85% Customer retention rate

13 Years
Top 10 clients relationship*

7 Years
Top 10 to 20 clients relationship*

* Years with BIP
(average on year 2018)

bip.

This docun

L'Economia del Corriere della Sera
Dir. Resp.: Luciano Fontana
Tiratura: 0 - Diffusione: 0 - Lettori: 2039000: da enti certificatori o autocertificati
13-LUG-2020
da pag. 21
foglio 1
Superficie: 77 %
CORRIERE DELLA SERA

Imprese

IL BUSINESS DELLA CONSULENZA

SHOPPING A LONDRA PER SBARCARE NEGLI USA COSÌ BIP SFIDA LA CRISI

La società con headquarter a Milano rileva la britannica Chaucer e punta a un fatturato di oltre 350 milioni. Lo Bianco, decano italiano del settore: «Operazione costruita mentre divampava la pandemia. Ma non ci siamo fermati» Adesso per il gruppo da tremila professionisti si apre il mercato anglosassone

di **Fabio Sottocornola**

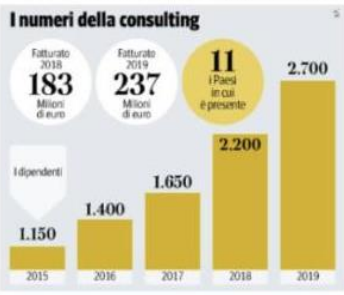
Colpo grosso nella consulenza. La Bip fondata a Milano da Nino Lo Bianco nel 2003 acquista a Londra la Chaucer management holding, prima tra le consulting britanniche interamente indipendenti. Un'operazione, che l'*Economia* racconta in anteprima e che vale almeno 60 milioni di euro, interamente per cassa. La più grande mai realizzata da Bip, oggi controllata al 61,5% dai francesi di Apax. Risultato? Si apprestano a nascere un colosso europeo della consulenza, che al termine di quest'anno realizzerà un giro d'affari di 350 milioni. Per Bip, un bel salto nel ranking mondiale: la ex Business integration partner lascerà la posizione numero 60 per entrare nelle prime 30. Ed è pronto, naturalmente, lo sbarco negli Usa.

Guidata da Chris Laslett, che resterà al comando delle operazioni nel primo periodo, Chaucer può contare su oltre 300 professionisti attivi nel seguire colossi dell'energia e dell'oil&gas come Bp, ma anche la pubblica amministrazione, dalla polizia agli enti della sicurezza nazionale fino al settore pharma. Le competenze digitali, maturate negli anni dalla società britannica, si sposano con quelle di Bip, da sempre forte nell'innovation management, nella security fino all'analisi dei dati. Almeno, questa è la speranza. Di più, l'orientamento di prospettiva che viene delineato dai vertici italiani: «Puntiamo a un'integrazione da realizzare in pochissimo tempo», racconta Lo Bianco, decano dei consulenti italiani e presidente di Bip, «vogliamo crescere come una One global company, per cultura, valori e approccio al mercato. Siamo molto attratti da questa operazione fatta con grande convinzione, non ci nascondiamo le difficoltà a integrare per la prima volta la cultura anglosassone nel nostro sistema globale».

Inutile dire che Lo Bianco, insieme ai soci e amministratori delegati Carlo Capè, responsabile per gli affari internazionali e Fabio Troiani, responsabile per il business in Italia, non sono preoccupati dagli scenari post Brexit. «Siamo abituati a lavorare con grandi aziende ed enti», spiega Lo Bianco, «non vendiamo prodotti ma servizi professionali in loco e siamo convinti che la consulenza crescerà ancora molto». Specialmente per l'ambito digitale. Infatti, lasciando da parte startup o nuove realtà che nascono con personale e vertici già orientati al virtuale, tutte le altre aziende devono attuare una trasformazione, che non sarà semplice e richiederà tempo.

Ma l'acquisizione inglese è stata un'occasione colta al volo o piuttosto un'operazione pianificata? Qui occorre fare un passo indietro. Da quando è entrato in maggioranza il fondo Apax France, la propensione internazionale per Bip è diventata una parola d'ordine: in particolare, nel 2019 hanno chiuso due operazioni. Dal Brasile, nel l'acquisto di Fdm, società con esperienza nel settore bancario, alla Spagna dove sono confluite nel gruppo altre 50 persone attraverso due spin off da Kpmg. Ma naturalmente rimaneva scoperto il fianco anglosassone e non solo. Su questo ha lavorato un apposito team di consulenti guidati da Andrea Alraghi con le classiche operazioni di scouting: almeno 70 i dossier visionati tra Regno Unito, ma pure Francia e Germania, che sono i Paesi a cui si guarda per il secondo semestre. L'obiettivo Chaucer è stato intercettato nel pieno di un'asta competitiva. Infatti, il fondo britannico Growth capital partners voleva uscire dal capitale. E qui, succede un fatto imprevisto: Bip si inserisce nella gara con una offerta economica che non viene considerata dal venditore «il miglior prezzo». Ma soprattutto, scoppia la pandemia da coronavirus. E tutto si ferma. Saltano almeno otto operazioni di m&a su dieci: nessuno si fida più a muoversi, i valori economici sono messi in discussione. «Noi abbiamo riflettuto molto su che cosa fare», spiega Capè, «e abbiamo deciso di proseguire nell'offerta». A far pendere verso Bip l'ago della bilancia sono stati gli stessi professionisti di Chaucer, «attratti da un percorso di lavoro che potremo fare insieme. Abbiamo avuto fiducia in noi stessi. Questo ci ha premiato».

Il piano di espansione si rivolge anche agli Usa dove Bip ha solo qualche ufficio. Molto più presente è Chaucer che sarà la testa di ponte per una presenza più massiccia. Peraltro, la consulenza pesa per l'1,2% del Pil negli Usa, per lo 0,8% nel Regno Unito e solo per lo 0,2% in Italia. «Anche se nel nostro Paese si è rivelato un business resiliente», sostiene Capè, «non abbiamo visto un grande calo in questi mesi». Il manager è convinto che la ripresa ci sarà: una classica ripartenza a V, di quelle molto veloci.



Volti Nino Lo Bianco, fondatore e presidente Bip. Sotto, Carlo Capè, amministratore delegato e responsabile per gli affari internazionali della consulting

A causa del Covid si è bloccato l'80% delle operazioni di m&a. Ma il settore si è dimostrato resiliente. E la ripartenza dell'economia sarà una veloce «V»

BIP Overview

Headcount | Global footprint

Our Offices:

Italy Milano, Roma,
Mogliano Veneto, Bologna
EMEA London, Madrid, Barcelona,
Bruxelles, Lugano, Wien, Zug,
Istanbul, Abu Dhabi
North America New York
South America São Paulo,
Rio de Janeiro, Santiago de Chile,
Bogotá

bip. xTech

Technology to Transform

Data

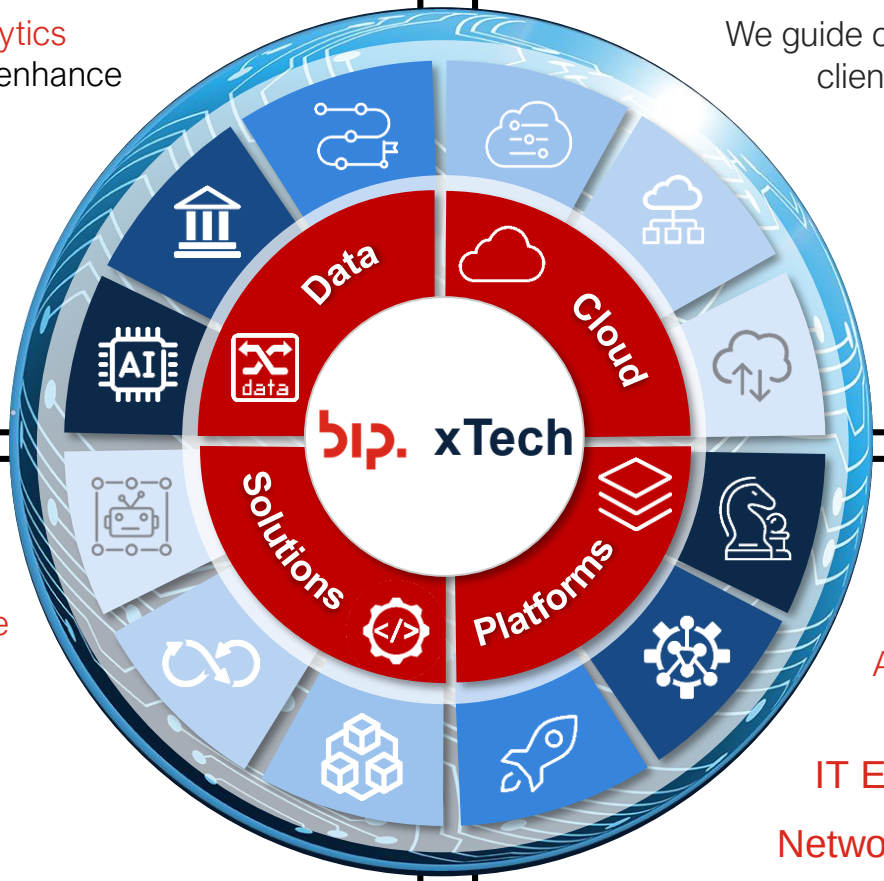
Our Domain Experts build next generation analytics solutions, setup data-driven organizations and enhance Data Management capabilities.

100+

Data Scientists

Data Visualizers

Data Governance Experts



Cloud

We guide our Clients through Cloud transition, helping our clients get the best value out of large transformation projects.

Cloud Data Architects

Cloud Data Engineers

50+

Machine Learning Engineers

Solutions

We build and deploy scalable, resilient and future-proof software solutions, leveraging Agile and DevOps methodologies and adopting modern architectures on cloud infrastructure.

RPA Specialist

50+

UI Developers

Full Stack Developers

Tech Platforms

We experiment with, evaluate and manage innovative technologies, such as Network Automation, 5G, Internet of Things, Blockchain, Augmented / Virtual Reality.

IT Experts / Architects

Network Experts

50+

Blockchain Experts

IoT Experts

bip. xTech

Certifications and Partnerships

Data Science

SAS, Microsoft Azure, Coursera, Data Science v2.0 – Dell Technologies Proven Professional, iGenius, neurala

Data Governance

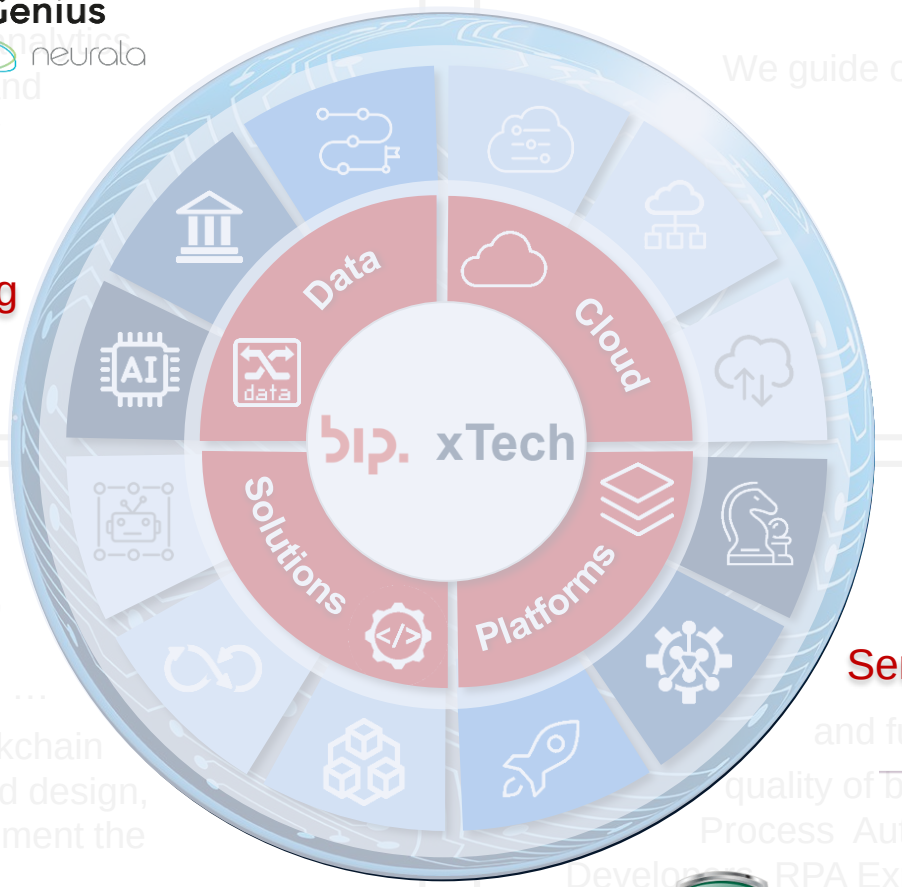
DAMA International, collibra, Informatica

Data Visualization

Qlik Sense, Power BI, tableau

Process Mining

celonis, myinvenio



Cloud Platforms

Google Cloud Certified, amazon web services, Microsoft Azure

Data Platforms

Google Cloud Certified – Professional Cloud Architect, Google Cloud Certified – Professional Data Engineer, amazon web services, Microsoft Azure, TERADATA, CCA cloudera SPARK & HADOOP DEVELOPER

Robotic Process Automation

UiPath Robotic Process Automation, blueprism, AUTOMATION ANYWHERE, NICE

Testing

Loadrunner

Programming

ORACLE Certified Professional Java SE 7 Programmer, LPIC-1

Devops

Microsoft Azure, amazon web services

Project Management

PROFESSIONAL SCRUM PSM MASTER, PRINCE2, PMP

Service Management

ITIL FOUNDATION, ISO 20000 Certified

IT Architectures

COBIT 2019 Foundation, TOGAF®

Networking

ZABBIX, tmforum CAREER CERTIFIED, ONF CERTIFIED SDN ASSOCIATE, CISCO CERTIFIED Architect, CISCO CERTIFIED CCNA, CISCO CERTIFIED CCNP

CRM Platforms

salesforce

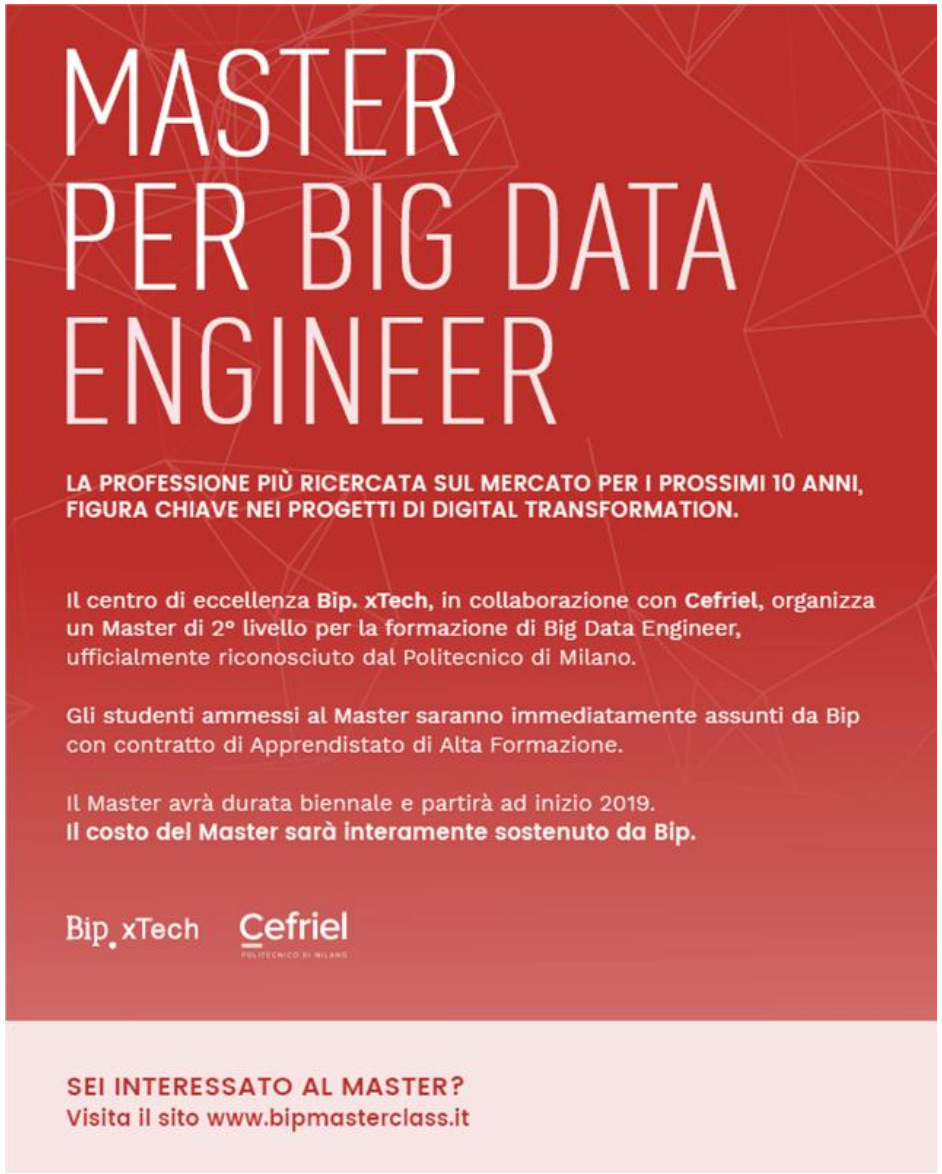
Second Level Master in Big Data Engineering

with Politecnico di Milano and Cefriel

Bip launched a new Second Level Master Class to train 20 Big Data engineer per session (productive after 5 months, duration of the master 2 years, direct hiring into Bip xTech workforce).

First class of trained Data Engineers available on projects since May 2019.

In 2020 a Cloud Data Architecture Master in launched to train 20 Cloud Data Architects.



xTech Data Analytics

Use cases

CUSTOMER ANALYTICS

- Churn Prediction
- Customer Experience
- Customer Journey
- Customer Segmentation
- Cross-Channel Analytics
- Speech Analytics

MARKETING INTELLIGENCE

- Marketing Mix Models
- Attribution Models
- Campaign Effectiveness
- Dynamic Pricing
- Up/Cross-selling Prediction
- Web Analytics

SMART PROCESSES

- Cognitive Helpdesk
- Autoresponders & Chatbots
- Automatic Issue Sorting
- Smart Repository
- Smart Document Search
- Process Mining

SMART PLANNING

- Sales Forecasting
- P&L Statement Forecasting
- Demand Planning
- Yield Prediction
- Traffic Prediction
- Agenda Management

RISK & CREDIT MANAGEMENT

- Collective Intelligence
- VAR and PAR
- Fraud Analysis & Reporting
- Risk Scoring

SUPPLY CHAIN ANALYTICS

- Supply Chain Optimization
- Workforce Planning
- Workforce Optimization
- On-time Delivery

PREDICTIVE MAINTENANCE

- Fault Detection
- Asset Management
- Anomaly Detection
- Root-Cause Analysis

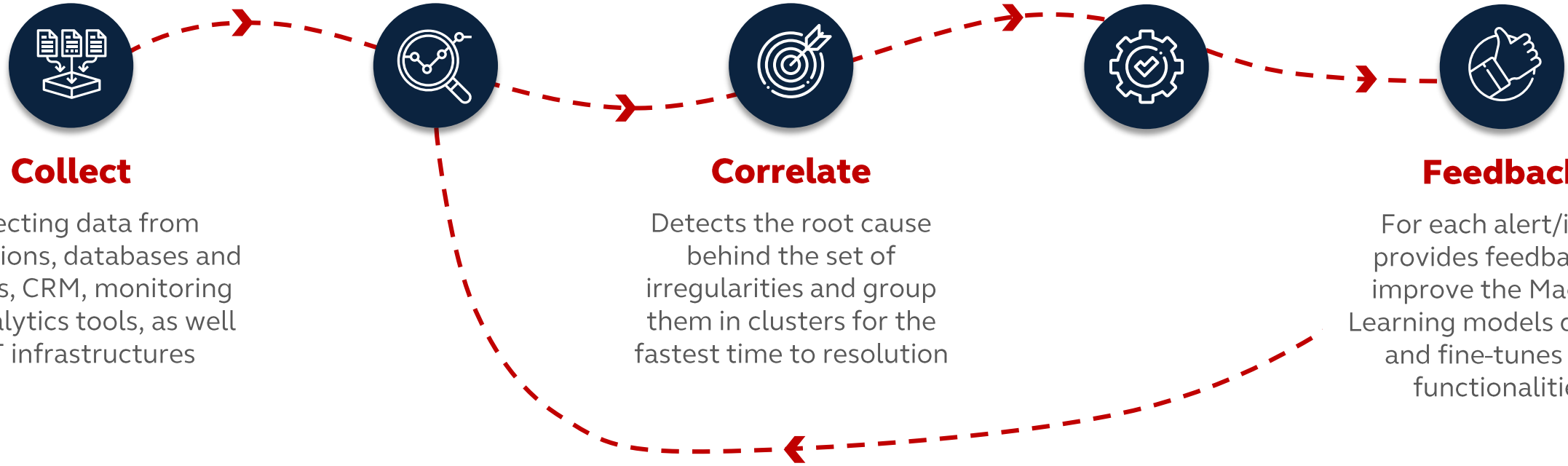
DATA MANAGEMENT & ARCHITECTURES

- Data Governance
- (Big) Data Platforms
- Data Architecture Migration
- Migration to the Cloud
- Data Platform Advisory

DATA STRATEGY & SERVICES

- Data Strategy
- Awareness Workshops
- Data Science Training
- Data Science Advisory
- Data Visualization
- Reporting & KPIs

Automated Data Quality Approach



Collect

Collecting data from applications, databases and streams, CRM, monitoring and analytics tools, as well as IT infrastructures

Analyze

Generating quality and business metrics based on history and normal behavior

Correlate

Detects the root cause behind the set of irregularities and group them in clusters for the fastest time to resolution

Act

Integrate the Data Quality solution in the existing workflow to address the issues quickly

Feedback

For each alert/issue provides feedback to improve the Machine Learning models deploys and fine-tunes tool functionalities

BIP Data Quality as a Service

DQaaS is a software solution developed by BIP, born from R&D activities (thesis) and industrialized on clients of different industries to provide automated data quality on their data

Fully customizable framework to address customer need and integrations

Can perform **analysis with few thousand datapoints**. Does not have an upper limit

Provides **data cleaning and fixing proposal** for low quality data and anomalies

Supports both **structured and unstructured data** as well as timeseries

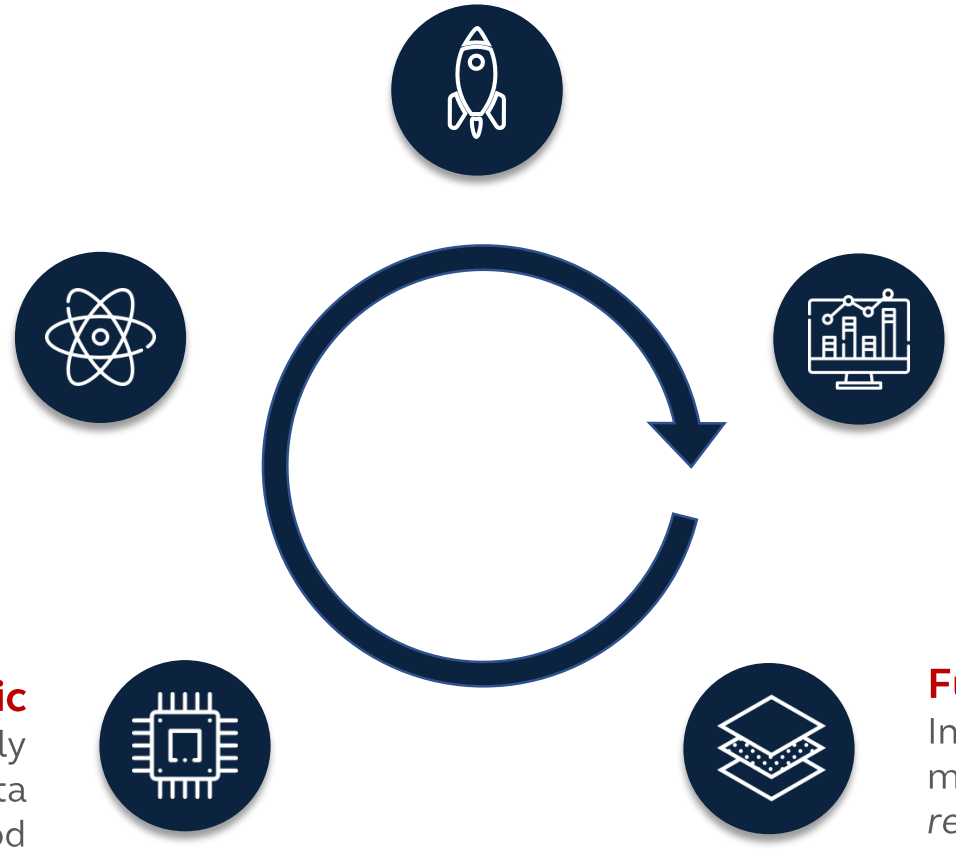
Handles **duplicated data as well as incorrect** or missing values

Can work with **Singular Events** (fine granularity) and **Aggregated Data** (e.g. number of visit per web page, number of transaction)

Key Features of BIP DQaaS

Auto-ML Engine

Autonomous generation and selection of multiple bots aimed at detecting and fixing quality issues, in diverse data environments without the requirement for domain knowledge



Automated Anomaly Detection

Autonomous data schema understanding, automated generation of quality checkpoints and false positive minimization

Feedback System

Continuous monitoring and reporting on issue identification. Users supervision and feedback input allow for validation and control

Data Domain Agnostic

Applicable to any dataset independently of domain, business sector and data storage method

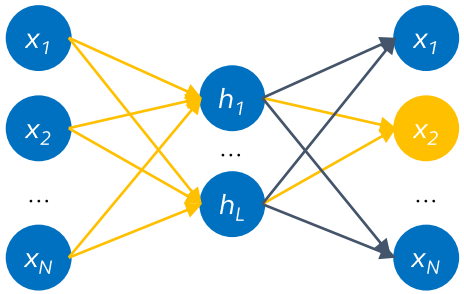
Full Data Coverage

Industrialized and scalable service model, on *data in transit* and *data at rest*. Compatible with both batch and streaming pipelines

Example of Probe Models

Several probe models are being implemented. Below is represented the working principle of three possible problems

Autoencoding



Description

- Autoencoders are deep neural networks that model their own input, in order to learn the relationship between the features, and map features to fewer dimensions in the hidden layer(s), to reconstruct the input

Anomaly detection

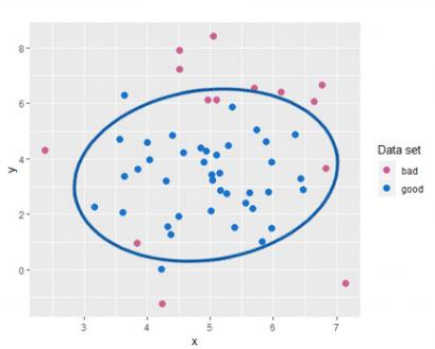
✓ If the predicted output differs from the input, the *DQ anomaly flag* is raised

Leave-One-Out

- Given *n* features (F) in a dataset, multiple models are built iteratively taking one feature at a time as target and leaving *n-1* features as predictors

✓ The predicted value is subtracted to the actual value. If the error is outside a statistical significance interval, the *DQ anomaly flag* is raised

Aggregated distribution



- Aggregated feature(s) distribution is inferred from values in the dataset

✓ If a feature for a given data does not belong to the relative inferred distribution, the *DQ anomaly flag* is raised

More than Data Quality

Due to DQaaS context-free probes, the applications are not limited to Data Quality issues

Data Quality Issues

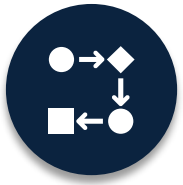
Automatically understanding the data, inference of Data Schema and potential Data Dictionary, detection of data quality and collection issues based on past data, provides potential fixing to null values and low-quality values. Visual interpretation of cluster of issues using Heatmap visualizer on the whole dataset.

Anomaly Detection and Root Cause Analysis

Detection of anomalies events/row inside the data not directly related to Data Quality Issues. Anomalies are detected when rows or events starts behaving in a not standard way or when the combination of metrics (features) are composed by outlier data. The anomalies are then grouped using a Clustering Engines to better understand the causes.

Data Drift Detection

Analysis and tracking of data trends during time (using temporal series and tabular dataset) to detect changes within past Dataset and alerting system. The solution can be used to adapt DQaaS models to new shapes of data or to trigger re-training of machine learning models working on the data analyzed.



Model Performance Monitoring

The framework can be used to monitor Machine Learning and Statistical model to control and prevent performance degradation providing feature dataset and prediction performed by the models. Similarly to anomaly detection and data drift, DQaaS highlights potential low-quality classifications and causes related to input data.

bip. xTech

Course Project 2020/2021

Rules and Dataset Explanation

HERE TO DARE

Course Project

Dataset Description

This year project is based on a **predictive maintenance** dataset for the prediction of faults on air conditioning equipment installed mobile network transmission sites in a 14-days forecast window. Available information are related to weather conditions (past and forecast), alarms and faults occurred on site, static features of the site.

air conditioning is crusial

- Daily data (April 2019 – Gen 2020 ~ 10 months) for each site
- Data are available for 2605 sites, distributed between training set (2071) and test set (534)

Variable/Variable pattern	Description	Type
SITE_ID	Unique identifier for the site belonging to the network	*int*
DATE	Reference date of the sample	*date*
N_TRANSPORTED_SITES	Number of neary sites for which the radio signal is transported through the site	*int*
CELL_TYPE_X	Indicates if the transmission cell of type X is mounted on site	*binary*
GEOGRAPHICAL_CLUSTER_K_x	Membership in the geographic cluster x (network clustered in 10 regions from 0 to 9)	*binary*
*mean/max/min_w*_prev*X*d	*Mean*, *max* or *min* of the weather condition *w* in the *previous X* days	*float*
*mean/max/min_w*_f_next*X*d	*Mean*, *max* or *min* of the forecasted weather condition *w* in the following *X* days	*float*
*cat*_sum_alarms_prev*X*d	Number of alarms associated to the category *cat* observed in previous *X* days. Alarms are classified in 9 categories. Details are available on attached excel	*int*
*cat_mean/max/min*_persistance_prev*X*d	*Mean*, *max* or *min* alarm duration (in minutes) of *cat* alarms in the previous *X* days	*float*
skew_*cat*_alarms_prev14d	Skewness indicator of *cat* alarms distribution in time in the previous X days	*float*
kurt_*cat*_alarms_prev14d	Kurtosis indicator of *cat* alarms distribution in time in the previous X days	*float*
aircon_sum_wo_target_next_14d	Binary target variable indicating the presence of a fault in the following 14 days	*binary*

Alarms categories

Course Project

Testing and Evaluation

Performances will be evaluated by means of the average daily **Weighted Recall** computed considering **the 10 sites with highest fault probability** as predicted with fault weighted by *N_TRANSPORTED_SITES* column. The thresholding level is therefore defined day by day by the top sites and not by a fixed threshold level.

Data Available in csv format: **https://we.tl/t-n6DOlLHf49** (archive password: *DMTMChallenge2020*)

Deadline: **Monday December 21 23:59**

We ask you to prepare and upload on Beep platform an archive containing:

- ***Prediction.csv***: you can find an ***example in the archive.*** The output of the prediction must be a ***fault probability value between 0 and 1***.
- ***Report.pdf***: ***4 pages*** to describe in detail your approach, data processing techniques, prediction model, performance computation methods and analytical results
- ***Presentation.pptx***: ***5 slides*** for the final project presentation describing you approach, data processing, prediction model (a summary of the report you wrote, imagine to present to a potential customer your approach)
- ***Scripts.zip***: any notebook or script you wrote will be evaluated, we expect ***a ordered list of script from data processing to prediction.csv file*** output.

The evaluation considers: Recall Score obtained on the Test Predictions, Report Quality, Presentation Speech, Quality of the code.

If you have any doubt you can ask directly on Beep forum.

THANK YOU

HERE TO DARE