

Data Mining and Text Mining Course Project

November 2020



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)



This document

1 di 1

L'Economia
del QUOTIDIANO DELLA SERA

Dir. Resp.: Luciano Fontana

Tiratura: 0 - Diffusione: 0 - Lettori: 2039000; da enti certificatori o autocertificati

CORRIERE DELLA SERA

13-LUG-2020

da pag. 21

foglio 1

Superficie: 77 %

Imprese

IL BUSINESS DELLA CONSULENZA

SHOPPING A LONDRA PER SBARCARNE 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

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, con l'acquisto di Fdm, società con esperienza nel settore bancario, alla Spagna dove sono confluite nel gruppo oltre 50 persone attraverso due spin off da

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

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



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»

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. Salvo 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 professionali 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.

© RIPRODUZIONE RISERVATA

BIP Overview

Headcount | Global footprint

Our Offices:

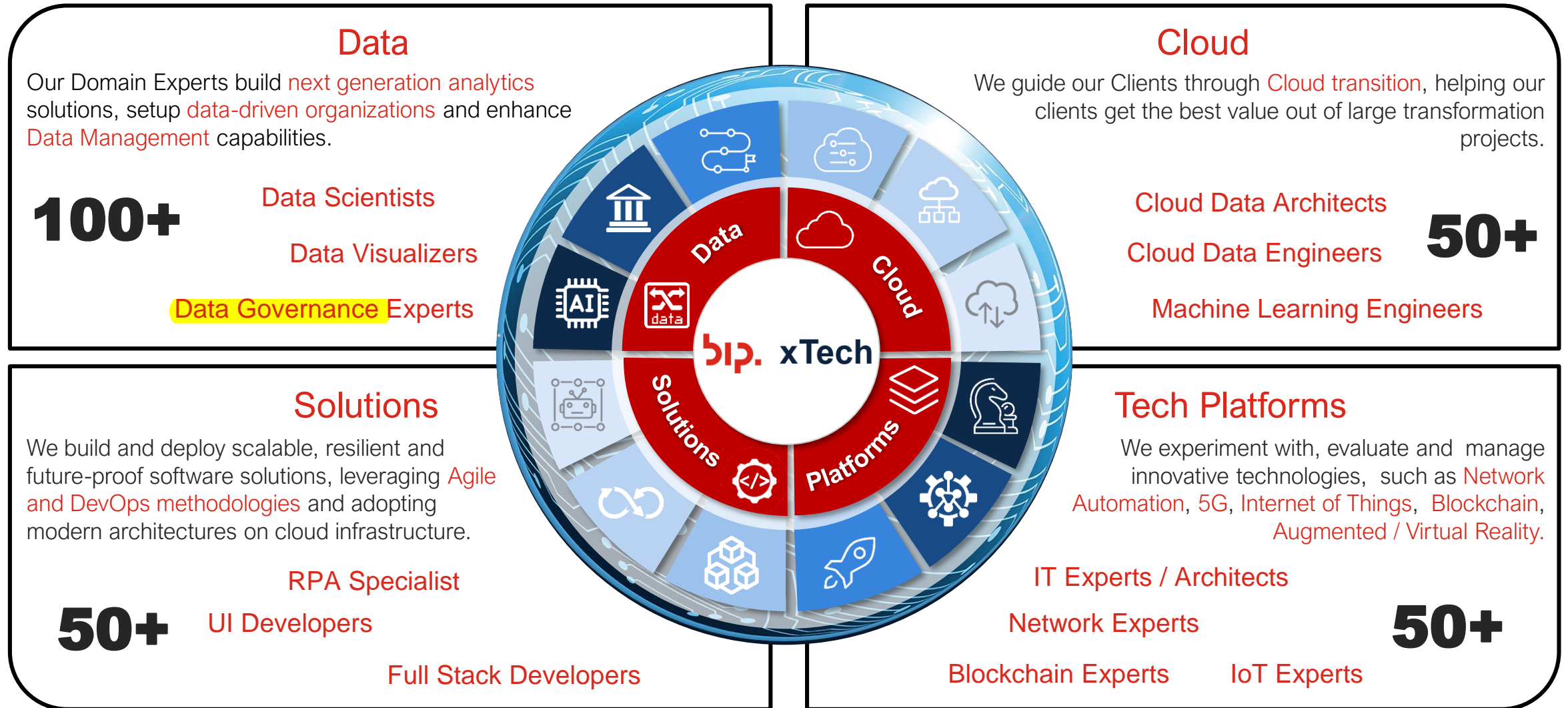
Italy Milano, Roma,
Mogliano Veneto, Bologna

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





Data Science



Data Governance



Data Visualization



Process Mining



Robotic Process Automation



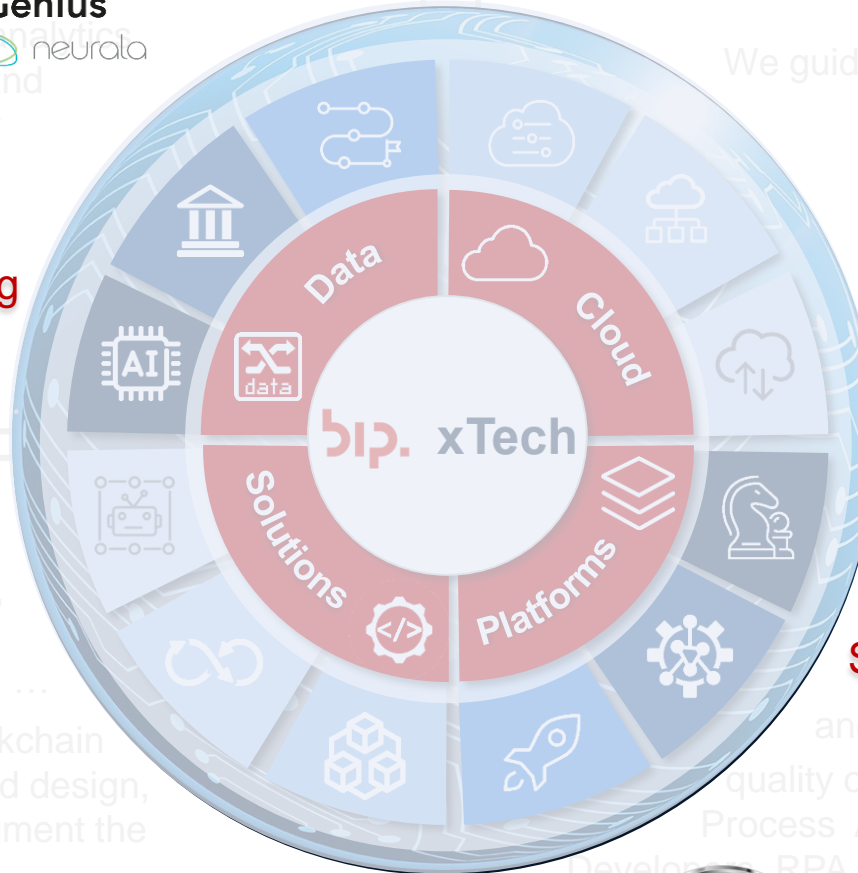
Testing



Programming



Devops



Cloud Platforms



Data Platforms



Project Management



Service Management



IT Architectures



Networking



CRM Platforms



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.



MASTER PER BIG DATA ENGINEER

LA PROFESSIONE PIÙ RICERCATA SUL MERCATO PER I PROSSIMI 10 ANNI,
FIGURA CHIAVE NEI PROGETTI DI DIGITAL TRANSFORMATION.

Il centro di eccellenza Bip. xTech, in collaborazione con Cefriel, organizza
un Master di 2° livello per la formazione di Big Data Engineer,
ufficialmente riconosciuto dal Politecnico di Milano.

Gli studenti ammessi al Master saranno immediatamente assunti da Bip
con contratto di Apprendistato di Alta Formazione.

Il Master avrà durata biennale e partirà ad inizio 2019.
Il costo del Master sarà interamente sostenuto da Bip.

Bip. xTech Cefriel
POLITECNICO DI MILANO

SEI INTERESSATO AL MASTER?
Visita il sito www.bipmasterclass.it

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

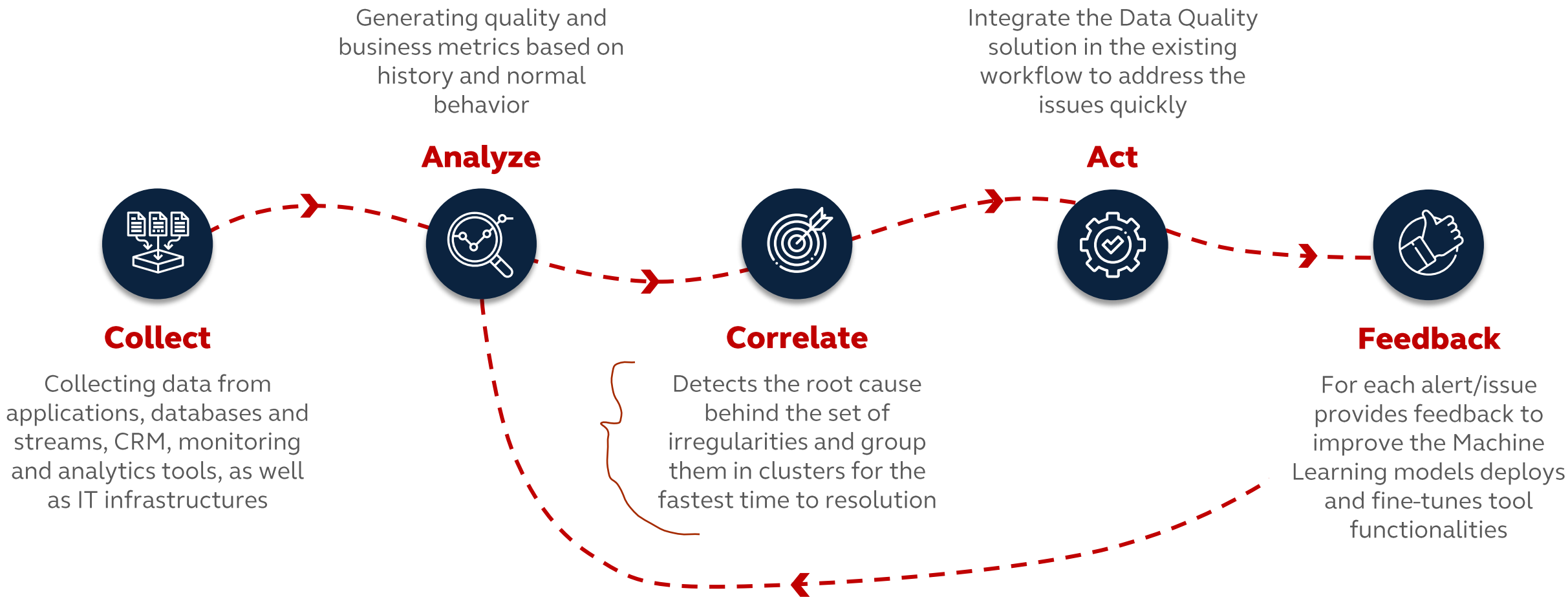


Data
Visualization



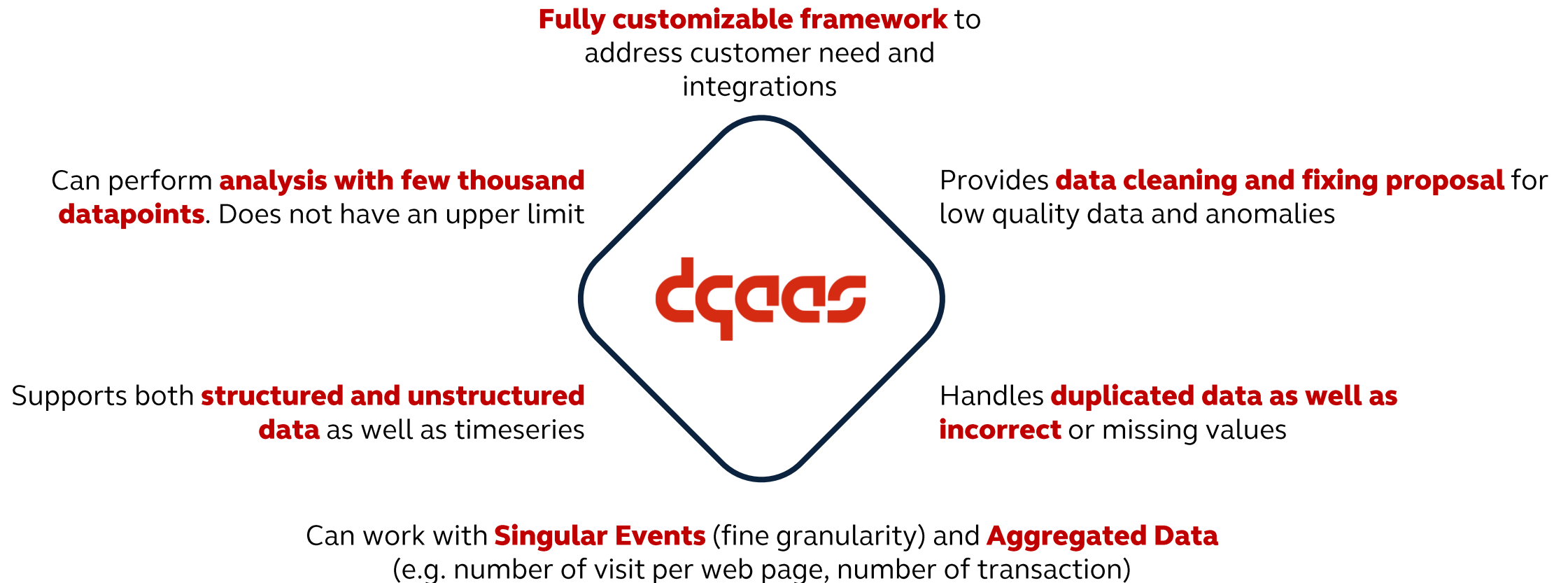
Reporting &
KPIs

Automated Data Quality Approach



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



Key Features of BIP DQaaS



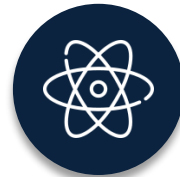
Auto-ML Engine

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



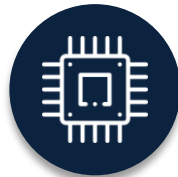
Automated Anomaly Detection

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



Data Domain Agnostic

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



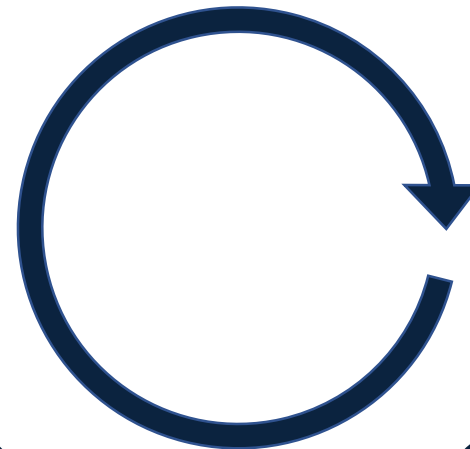
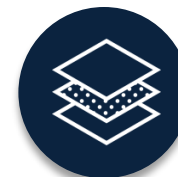
Feedback System

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



Full Data Coverage

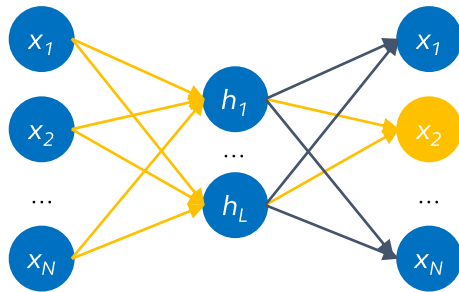
Industrialized and scalable service model, on *data in transit* and *data at rest*. Compatible with both batch and streaming pipelines



Example of Probe Models

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

Autoencoding



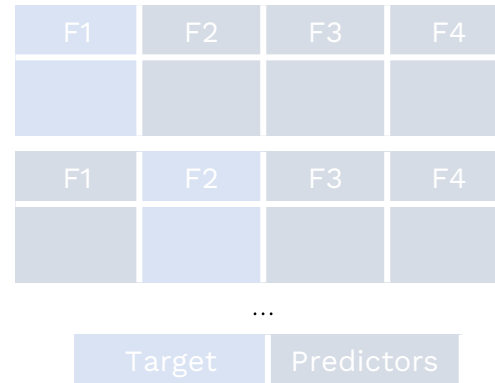
Description

- 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

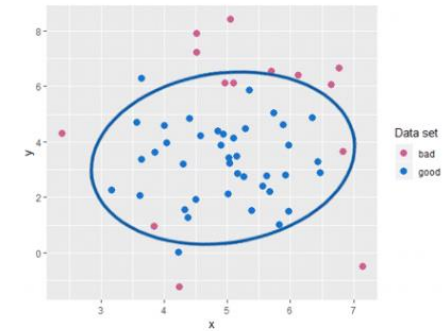
- ✓ If the predicted output differs from the input, the *DQ anomaly flag* is raised

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.



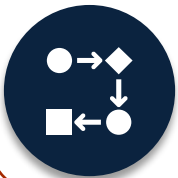
Anomaly Detection and Root Cause Analysis

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.

Course Project 2020/2021

Rules and Dataset Explanation



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.

- 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	<u>int</u>
DATE	Reference date of the sample	<u>date</u>
N_TRANSPORTED_SITES	Number of neary sites for which the radio signal is transported through the site	<u>int</u>
CELL_TYPE_X	Indicates if the transmission cell of type X is mounted on site	<u>binary</u>
GEOGRAPHICAL_CLUSTER_K_x	Membership in the geographic cluster x (network clustered in 10 regions from 0 to 9)	<u>binary</u>
mean/max/min_w_prevXd	Mean, max or min of the weather condition w in the <u>previous X days</u>	<u>float</u>
mean/max/min_w_f_nextXd	Mean, max or min of the forecasted weather condition w in the <u>following X days</u>	<u>float</u>
cat_sum_alarms_prevXd	Number of alarms associated to the category cat observed in previous X days. <u>Alarms are classified in 9 categories.</u> Details are available on attached excel	<u>int</u>
cat_mean/max/min_persistence_prevXd	Mean, max or min alarm duration (in minutes) of cat alarms in the previous X days	<u>float</u>
skew_cat_alarms_prev14d	Skewness indicator of cat alarms distribution in time in the previous X days	<u>float</u>
kurt_cat_alarms_prev14d	Kurtosis indicator of cat alarms distribution in time in the previous X days	<u>float</u>
aircon_sum_wo_target_next_14d	Binary target variable indicating the presence of a fault in the following 14 days	<u>binary</u>



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-n6DOILHf49> (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 your 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

