

Data Mining and Text Mining Course Project

May 2020

HERE TO DARE



BIP Overview

BIP in a “nutshell”



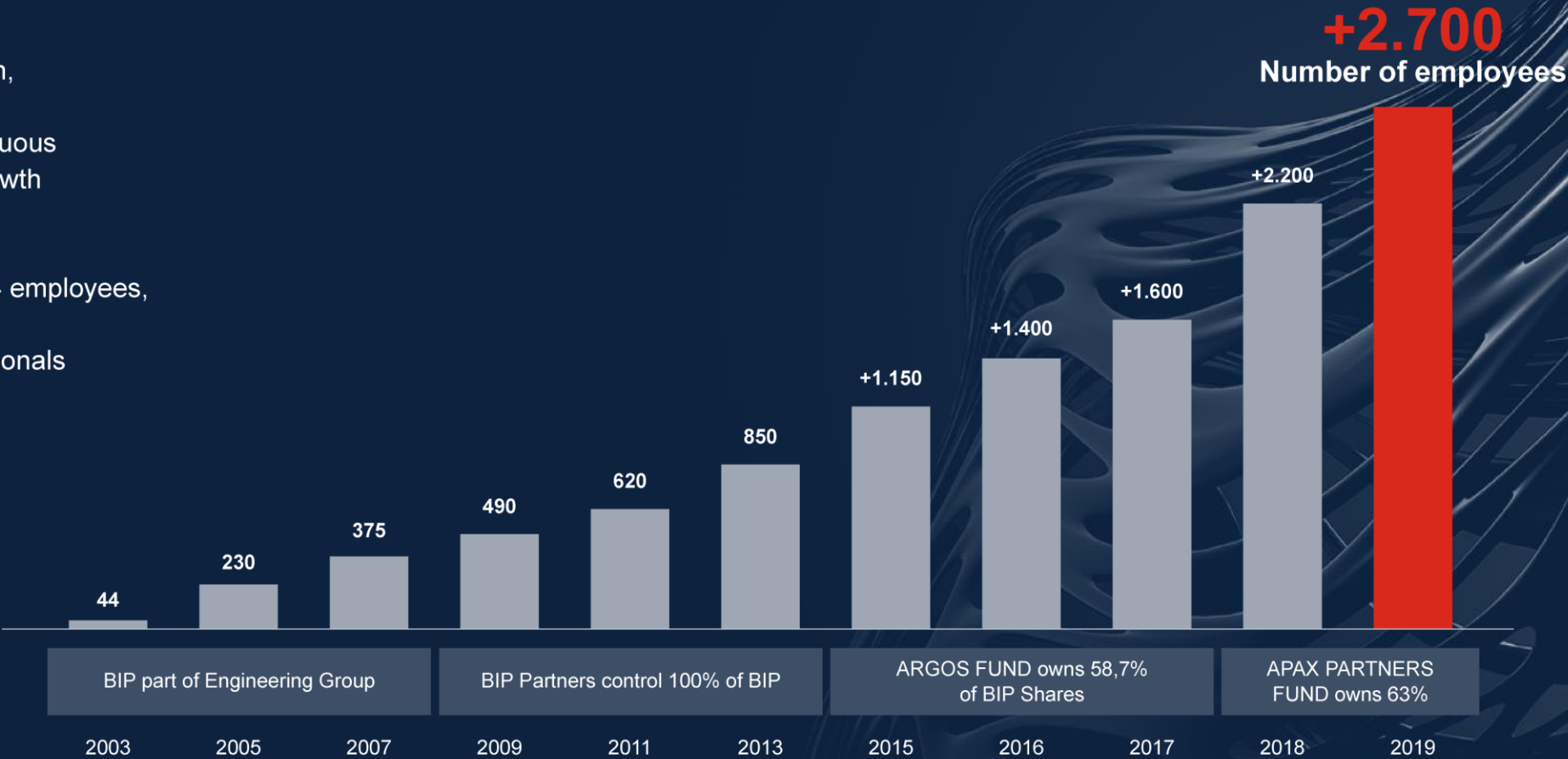
* Years with BIP
(average on year 2018)

BIP Overview

Headcount | Timeline and major growth steps

01
Since its foundation, BIP has always experienced continuous and impressive growth

02
Starting out with 44 employees, today BIP employs over 2.700 professionals



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

Our Domain Experts build **next generation analytics** solutions, setup **data-driven organizations** and enhance Data Management capabilities. By leveraging one of the largest, professional European Data Scientist communities - tightly coupled with **Data Governance** Experts and **Data Strategists** - we address business challenges getting the most out of data through analytics.

Cloud

We guide our Clients through **Cloud transition**, helping our clients get the best value out of large transformation projects. Our Cloud Architects and Engineers deliver efficient multi-cloud governance, optimized **applications and data architectures**, and effective Cloud migration approaches.

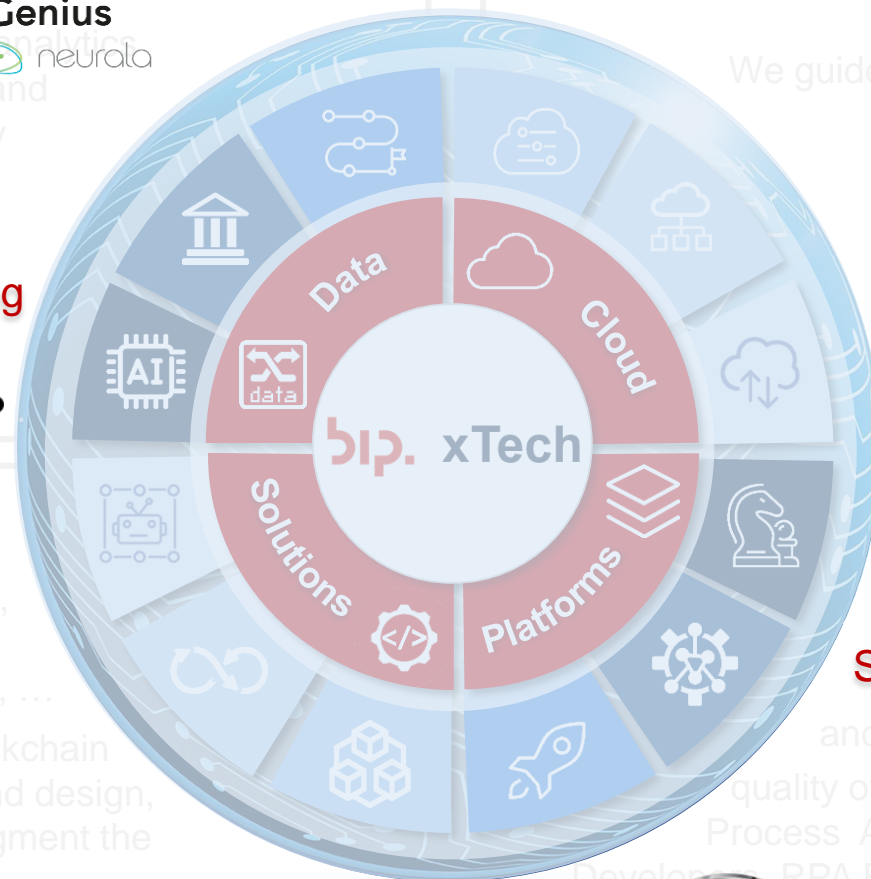
Solutions

We build and deploy software solutions, leveraging **Agile and DevOps methodologies** to speed up and tailor-fit the result, adopting modern architectures on cloud infrastructure to realize scalable, resilient and future-proof platforms. Improve efficiency and quality of business processes developing **RPA** (Robotic Process Automation) bots. Our team includes Full-Stack Developers, RPA Experts, DevOps engineers and Agile Coaches

Platforms

We experiment with, evaluate and manage innovative technologies, such as **Network Automation** (Virtualization & Orchestration), **5G**, Internet of Things (**IoT**), **Blockchain**, Augmented Reality/Virtual Reality (**AR/VR**), ... Our IT Architects, Network Engineers, Blockchain and AR/VR Experts help define strategy and design, implement these solutions and more to augment the enterprise's digital foundation.





Data Science



Data Governance



Data Visualization



Process Mining



Robotic Process Automation



Testing



Programming



Devops



Cloud Platforms



Data Platforms



Project Management



Service Management



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.

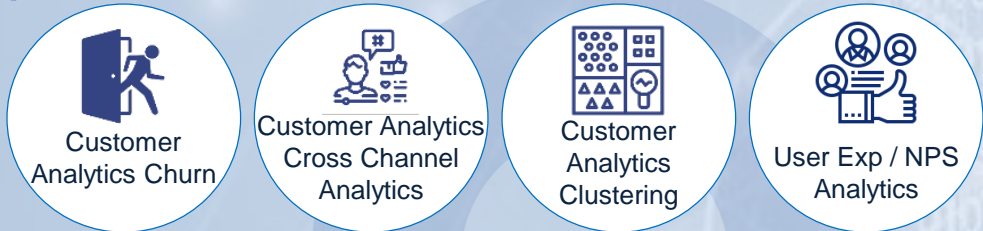
Gli studenti ammessi al Master saranno immediatamente assunti da Bip
con contratto di Apprendistato di Alta Formazione.

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

Customer Behaviour / Experience Analytics



Marketing and Campaign Management



Process analytics and automation



Cognitive Applications with Text, Audio, Video



Sales and Stores



Supply Chain and Logistics



Finance and Cash



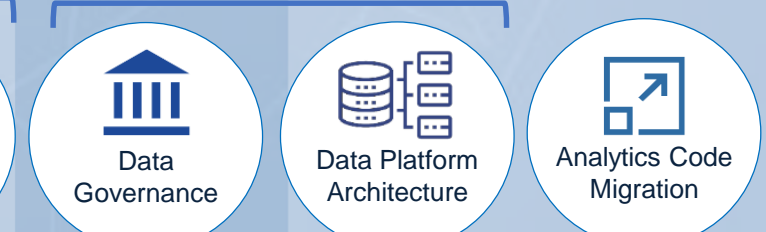
Analytics on Technology



Other Use Cases



Data Platforms Design and Data Management



Delivered Gas Forecasting

December 16, 2019

Background

Natural gas is the most important energy source in Italy, fueling domestic heating, industrial facilities, and thermoelectric power plants. The transport of natural gas is guaranteed by a network of over 32 thousand kilometers of pipelines, with the activity entrusted to the transmission system operator (TSO). According to government regulations, **the TSO must also provide gas demand forecasts to ensure safe, reliable and efficient operational planning** and drive gas prices as well as future investment requirements. This task results particularly difficult in Italy due to the complicated structure of the gas network as well as volatile weather patterns.

The TSO initially generated forecasts using time series models. However, due to the increasing complexity of the gas network infrastructure and protocols, it became apparent that a new and more advanced approach was required.

Decisions and Actions

Our proposal was to develop **a forecasting system capable of increasing the accuracy and stability of Italian gas demand forecasts through a Machine Learning approach**. The main focus was on the one-day-ahead model, subject to an economic incentive proportional to the daily percentage error.

The system:

- provides hourly predictions of the volume of gas delivered across Italy for the following 4 days, with detail per type of usage (Thermoelectric, Industrial and Civilian);
- utilizes a set of neural network models that consider real-time gas usage, weather forecasts, historical Company data and electricity demand forecasts;
- has a rolling operating mode, with the models continuously updated with the most recent data.



Results:

better network balancing, increased efficiency of gas transportation activities

improvement of the quality of the information about the imbalance of the system, leading to more accurate market interventions

+40% *performance boost with the day-ahead model achieving a 2.9% yearly average percentage error and significant economic subsidies.*



Course Project 2020

Rules and Dataset Explanation

HERE TO DARE



Course Project

Dataset Description

This year project is based on a **sales forecasting** dataset with aggregated information for different products (SKU) throughout a 3-years time window:

- Weekly data (Dec 2016 – Dec 2019)
- Data are available for 43 SKU but the target for the prediction is restricted to 12 SKUs

Variable	Description	Type
SKU	Unique identifier for the products	<u>int</u>
Pack	Type of pack in which the product is sold	<u>str</u>
Size (GM)	Product weight	<u>float</u>
Brand	Product brand	<u>str</u>
Price	Planned price of sale for the product in week w	<u>float</u>
POS_exposed w-1	Number of stores in which the product was put on evidence at w-1	<u>int</u>
Volume_on_promo w-1	% Volume of product put on promo at w-1	<u>float</u>
Sales w-1	Sales of product at w-1 (lagged target)	<u>int</u>
Scope	Boolean that indicates SKUs in scope (target)	<u>bool</u>
Target	Sales of product in w	<u>int</u>

Course Project

Testing and Evaluation

Performances will be evaluated by means of the **MAPE (Mean Absolute Percentage Error)** metric on the test set, which is provided net of the target column.

Data are split time-wise into train and test sets with the following criterion:

- 2 years and 6 months in the **train set (Dec 2016 – July 2019)**
- 6 months in the **test set (July 2019 – Dec 2019)**

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right|$$

A_i : actual value
 F_i : forecast value

Data Available in csv format: <https://we.tl/t-2zID8BelrK> (archive password: *DMTMChallenge2020*)

Deadline: **Friday June 12 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
- **Report.pdf**: 4 pages to describe your approach, data processing, prediction model and analytical results
- **Presentation.pptx**: 5 slides for the final project presentation describing you approach, data processing, prediction model
- **Scripts.zip**: any notebook or script you wrote will be evaluated

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

THANK YOU

HERE TO DARE

