

About Us

Born in 2003, we have collected the historical experience of consulting and have added two key ingredients: **innovation and digitalisation**.

Thanks to this path we today have acquired over **4.000** professionals, present in countries, covering over **4.500** projects, owning the most advanced knowledge in the field of Digital Transformation, Data Science, Cybersecurity, Industry 4.0, IOT and of all disruptive technologies that we put at the service of every market sector.

Why BIP.xTech - The largest professional Data Scientist community in Italy

xTech's international community of curious colleagues, passionate about data, AI, cloud and exponential technologies, have a shared purpose, to make a positive impact – for our clients, ourselves, and our communities.

Internship Topic

1. Bond risk rating classification

- Context: in investment, the bond credit rating represents the credit worthiness of corporate or government bonds. It is not the same as an individual's credit score. The ratings are published by credit rating agencies and used by investment professionals to assess the likelihood the debt will be repaid. Credit rating is a highly concentrated industry with the "Big Three" credit rating agencies – Fitch Ratings, Moody's and Standard & Poor's (S&P) – controlling approximately 95% of the ratings business. While government bond are typically largely covered by credit agencies, many corporate bonds remain unrated. Due to lack of information, unrated corporate bonds will offer typically higher returns and represent a good opportunity for investors if they are able to accurately identify those with low risk.
- Thesis objective: the aim of the thesis is to build an effective Machine Learning classification engine to predict corporate bonds rating based on past rating data.

2. AI-driven market monitoring

- Context: while monitoring known important market aspects (e.g., key regulators, competitors, influencers, etc) is important to many as well as recognized and performed by all companies, monitoring "the unknown" still remains a big challenge for most enterprises and often left to human judgment. Being able to detect early trends, shift in seasonality, anomalies, new to topics and more is especially key for capital markets players to better manage risk and increase return opportunities.
- Thesis objective: the goal of the thesis is to build a Business Intelligence monitoring tool (refreshed daily) to analyse news and/or social data. The system will include: (1) data gathering pipelines, (2) backend databases, (3) ML engines to detect topics (unsupervised learning), (4) time series analysis engine to detect seasonality, anomalies, trends, anomalies, and (5) a Business Intelligence dashboard to visualize and analyse data.

3. AI-driven Search Engine result explanation

- Context: modern search engines working on unstructured data rely on machine learning and vector representation of the information (e.g., BERT, GPT-2/3) to provide fewer static results than a keyword-based approach. However, the interpretability of the information ranking lacks a human-understandable explanation which leads to a preference for a standard keyword-based system.

- Thesis objective: the goal is to analyze and study an AI-driven search engine on real-world data (legal documentation and open data sources) that helps users to understand the retrieved information based on their content.

4. Question-Answering Pair Generation from technical documentation

- Context: Maintaining FAQs, Chatbot Datasets or operator training material requires a lot of human effort to keep them up to date and typically has a lower priority than the generation and creation of updated documentation (e.g., technical documents, commercial material, usage documents), new generative models allow the generation of question and answering pairs starting from a big and general documental base.
- Thesis objective: usage of generative models (also pre-trained) to create an automatized system that maintains such question-answering pairs when new documents are available based on a specific set of keywords or product list

5. Synthetic Data Generation

- Context: Synthetic data is artificial data generated with the purpose of preserving privacy, testing systems or creating training data for machine learning algorithms. Synthetic data is generated using different algorithms that mirror the statistical properties of the original data but does not reveal any information regarding real people. Industry leaders started to discuss the importance of data-centric approaches to AI/ML model development, to which synthetic data can add significant value. Especially when companies need data to train machine learning algorithms and their training data is highly imbalanced (e.g. more than 99% instances belong to one class), synthetic data generation can help build accurate machine learning models.
- Thesis objective: testing and benchmarking generative data algorithms to evaluate gains in predictive capabilities of Machine Learning models

6. Online Newspapers' Dynamic Paywall Optimization using Reinforcement Learning

- Context: One of the most important subscription channels for online newspapers is the paywall, which is a "registration wall" that blocks access and asks readers to create an account and subscribe to an offer. An optimal choice of "when" the paywall should be shown (e.g. when the reader is more engaged while reading), "what" the paywall should offer (e.g. personalized subscription offers, more free content, advertising), and "how" it should render its content (e.g. showing a preview of the article before showing the paywall) has a relevant impact in the efficacy of the paywall, increasing readers conversion rate, subscription volumes and the average value of the client.
- Thesis objective: design, implement and test a Machine Learning model that drives the dynamic paywall of an online publisher, learning the optimal strategy online, adapting and evolving the strategy over time by getting new data and outcomes of tested strategies.

7. Explainable Deep Reinforcement Learning in Multi-Agent models

- Context: Explainable RL is a set of methods aiming to provide explanations so that human users have a better understanding of the RL agent's behavior. This is even more complicated when facing Deep RL. The objective of the thesis is to explore and summarize current policies and provide suggestions to increase explainability. A particular focus should be on the possibility to perform explainability on MADRL models, by considering the contribution of each agent in achieving a collaborative goal.
- Thesis objective: the goal would be to detect agents with important contributions and understand whether there are conflicts.

8. Derive relationships among attributes and business rules in data by exploiting AI explainability algorithms

- Context: AI explainability can be exploited to understand constraints and relationships among data attributes. This kind of information should be provided by domain experts through a data glossary, but in the real world, data scientists have to perform analysis on their own to understand the data
- Thesis objective: create a solution that automatically derives relationships and rules among data attributes. A particular focus should be on the possibility to generate a set of constraints that can be employed as acceptance rules for new incoming data