



# GREENCHAIN ANALYTICS

**Pioneering Sustainable Intelligence for  
Tomorrow's Supply Chain**

A Supply Chain-Focused Startup Integrating an Eco-Friendly Approach into Its Solutions

## The team :

- A **technical team** composed of four data scientists, dedicated to develop robust prediction models able to handle complex or incomplete data.
- A **business team** of two, focused on understanding client needs and providing real-time progress updates on technical advancements to our partners.

## The mission :

- Schneider Electric approaches us with a dataset comprising 2.2 million entries, accompanied by other sub-datasets of various indicators.
- Our goal: To **predict the sales volume** for the fourth month of each third. In line with our principles, we also aim to propose measures to **reduce the carbon footprint** across the entire supply chain.

## The challenge :

- Responded immediately to the proposed challenge, mobilizing all available resources.
- In this context, the technical team collaborated, sharing databases to promptly and most importantly prepare the dataset thoroughly, especially in handling missing data. The goal was to establish a predictive model as accurate as possible within the given timeframe.

# Project management



## Understanding and processing the data

- **Analyze the distribution** of the databases to understand them.
- Operate the relevant **formatting** on the features. For instance, the life cycle column requires the NaN values to be replaced by "active".
- **Merge** the extra datasets to add quantitative data.
- **Remove irrelevant data.** For example, some columns are too specific and can be substituted for broader ones. For instance remove Zone and Country and use Region instead (contains less categories).
- **One Hot Encoding** of the few categorical features left, as the product division, to increase model efficiency.

## Modeling and Deployment:

- We want to forecast a sales volume which is a **continuous quantity**. As a consequence, we will use a regression model.
- To conveniently try a variety of models, we **standardize** the data.
- Amongst the different models tried, we chose to use **XGBRegressor** from the XGBoost library (tree-based algorithm).
- We use a **Bayesian search** to identify the best hyperparameters.
- We then train our model (being mindful of over-fitting) and use it to predict the test sample.

# Sales Forecasting and Sustainability

Focusing on immediate efficiency and environmentally friendly approaches, we present various perspectives to Schneider Electric and would be thankful to initiate more collaboration on sustainable solutions.



## Renewable Energy Production Forecasting

- Employ data mining to forecast the generation of renewable energies such as solar or wind, and **adjust production schedules** based on the availability of renewable energies, ensuring a more **energy-efficient operation**.

## Optimized Production Schedules:

- By adjusting production based on sales forecasts, we can effectively align output with actual demand. This strategy not only optimizes resource utilization but also significantly reduces energy consumption by **minimizing unnecessary machine operation times**.
- This strategy also helps **prevent overproduction**, thereby reducing resource wastage and energy usage.

## Failure Equipment Forecasting:

- Use data mining to predict equipment failures, allowing for **proactive maintenance interventions**. This approach not only reduces unnecessary downtimes but also saves energy.
- Schedule maintenance activities based on predictive data to minimize disruptions and **enhance energy efficiency**.





**We have to stay on top of the upcoming environmental challenges by optimizing business models.**

**One of the key aspects of this is to to predict sales volume. This is why we have to continuously strive to improve our models : their efficiency, robustness and their relevance as regards to the environment and associated legal norms.**