Track **UAV**

🏷️ Fault Detection and Preventive Maintenance of Drones

🧐 Context

The use of drones in increasingly diverse fields (logistics, surveillance, infrastructure inspection, or even rescue operations) comes with growing challenges of reliability and safety.
Drone failures can lead to **loss of equipment**, **service interruptions**, or **risks to people**.
It is therefore essential to prevent breakdowns through **preventive maintenance** and **early fault detection** approaches, in order to intervene before an incident occurs or to quickly identify its cause.

🎯 Objective

Based on the provided sensor data (in full or in part), the objective is to develop an analysis and anomaly detection model capable of, depending on the chosen approach:

- **Detecting the presence of a fault** in the drone's operation;
- **Identifying the probable origin** of the fault;
- **Assessing the severity** of the detected fault to guide maintenance actions.

In short, the goal is to implement a **predictive maintenance** approach to improve the availability, safety, and lifespan of drones.

📚 Provided Dataset

- **Format:** Matlab files easily usable or convertible in Python (via scipy.io.loadmat)
- **Link:** DronePropA: Motion Trajectories Dataset for Commercial Drones with Defective Propellers **(Article)**
- **Volume:** 130 files in total across 4 main categories — it is sufficient to use only part of them.
- **Quick Description:**
  This dataset is one of the few presenting real operational data from drones with and without faults.
  Conducted indoors under controlled conditions, the tests clearly recorded sensor data from the drone and its controller over several paths ("trajectories") and multiple drone states (healthy and 4 fault types).

Each file describes the sensors of both the controller and the drone, with the file name indicating the drone's state (F0 = no fault, t2 = trajectory 2, etc.).

🔧 Expected Deliverables (NON-MODIFIABLE)

- Project description + code (GitHub repository)
- Pitch deck (.pdf)
- Demo video (≤ 5 min)

👩‍💼 Expert Contact Point

- **Name:** Sarah Ounes
- **Position:** President Davinci Hive
- **Email :** sarah.ounes@edu.devinci.fr

Track **Finance**

🏷️ Detection of Fraudulent Bank Transactions

🥸Context

Financial institutions are facing increasingly sophisticated fraud attempts on cards and online payments.
IBM provides a realistic dataset covering **2016–2018** (around **210,000 transactions**) to train and evaluate fraud detection models.
A key feature: **fraudulent clients present in the training set are different from those in the evaluation set.**
The models must therefore **generalize to new clients (cold start)**, reflecting a real-world operational scenario.

🎯 Objective

- Design, train, and document a Machine Learning model capable of **predicting whether a transaction is fraudulent (1) or not (0)**.
- **Produce a submission file** in the required format containing transaction_id and fraud_prediction for each transaction in the evaluation set.
- Follow the constraints: **Do not use evaluation_features.csv for training** (it is for inference/evaluation only).
- Take into account **generalization to new clients** and classical challenges (class imbalance, missing data, temporal validation, etc.).
- Create a **visual dashboard**

📚 Provided Dataset

- **Source:** hackthon data.zip
- **Format:** CSV
- **Volume:** ~210,000 transactions (years **2016–2018**)

Files:

- transactions_train.csv: training transactions (features).
- train_fraud_labels.json: labels (fraud / non-fraud) for training.
- cards_data.csv: information on payment cards.
- users_data.csv: user information.
- mcc_codes.json: MCC codes (merchant categories) and descriptions.
- evaluation_features.csv: evaluation transactions without labels (to be used only for prediction).

Quick Description:

- Transactional variables (amounts, dates, merchants via MCC, etc.), enriched with **card** and **user profile** data.
- Fraud labels provided separately for training.
- **Challenge:** Fraudulent clients are **not overlapping** between training and evaluation → **transfer/generalization capability required**.

🔧 Expected Deliverables (NON-MODIFIABLE)

- Project description + code (GitHub repository)
- Pitch deck (.pdf)
- Demo video (≤ 5 min)

👩‍💼 Expert Contact Point

- **Name:** Mehdi Boulaymen
- **Position:** AI Engineer
- **Email :** Mehdi.Boulaymen@ibm.com

DEVINCI JUNIOR
Paris - Nantes

Track **Sustainability**

🏷️ Estimation and Simulation of the CO_2 Impact of LLM (Large Language Model) Queries

Capgemini engineering

🤔 Context

Large language models (LLMs) such as **LLaMA**, **Falcon**, and **Mistral** are widely used for text generation and task automation.

Their increasing use raises **environmental concerns**, particularly regarding **energy consumption** and the **carbon footprint** associated with each inference request.
This use case aims to **understand, model, and simulate** this impact based on real data from open-source models, with the goal of promoting **more sustainable AI**.

🎯 Objective

Design a **complete solution combining a user interface and an AI model** (Machine Learning or Deep Learning) capable of **estimating the carbon footprint (CO_2e)** of a request sent to an LLM.
Students will need to:

- **Analyze** the energy consumption of an open-source model.
- **Build** a predictive model estimating the energy consumed per **token** or per **request**.
- **Convert** this energy into a **CO_2e equivalent (kg)**, depending on the energy mix used.
- **Develop an interactive interface** (web app, dashboard, or browser extension like Chrome) to **visualize and simulate the environmental impact** of a query.

💡 **Bonus:** Extend the simulation to a **complete conversation** and/or **compare the environmental impacts** across multiple LLM models.

📊 Dataset

Source : Hugging Face – ejhusom/llm-inference-energy-consumption
Format: CSV
Volume: About 5,200 rows, 80 variables, across 15 configurations (models and different hardware setups).

Description:
This dataset contains real measurements of **energy consumption** during inference from several open-source language models (**LLaMA-3**, **Gemma**, **CodeLLaMA**, **Falcon**) on different hardware types (**laptop, workstation, server**).

Key variables include:
model_name, hardware_type, power_draw, inference_time, energy_consumption_llm_total, prompt_token_length, response_token_length, word_count, etc.
Note:
The dataset does **not** include proprietary models like GPT-4, Gemini, or Claude. However, it is possible to **extrapolate** their energy consumption based on open-source models (e.g., LLaMA-2-70B or LLaMA-3-70B) by applying **scaling factors**.

🔧 Expected Deliverables (NON-MODIFIABLE)

- Project description + code (GitHub repository)
- Pitch deck (.pdf)
- Demo video (≤ 5 min)

👩‍💼 Expert Contact Point

- **Name**: Hernan Carrillo
- **Position** : Data & Ai Scientist
- **Email** : hernan-camilo.carrillo-lindado@capgemini.com

DEVINCI JUNIOR
Paris - Nantes

Track **Education/Automation**

🏷️ Intelligent Help Center for PLV Students

DE VINCI

🧐Context

The current Help Center relies on **manual searches** in a knowledge base containing about **400 questions, answers, and tutorials** for students.
This method is inefficient and does not allow for **natural interaction** or **contextual analysis** of requests.
The goal is therefore to improve the user experience by developing an **intelligent conversational agent** capable of understanding the context of a question and providing the most relevant answer — or, if necessary, **redirecting to the correct channel** (form or email).

🎯 Objectif

Develop a **Proof of Concept (POC)** for an intelligent conversational agent that:

- **Ingests** and uses an Excel file containing questions, answers, topics, and metadata.
- Provides **enriched HTML-formatted responses** through a web or mobile interface.
- Analyzes the **context of the query** and proposes the **best possible answer**.
- Automatically **redirects** to a **form or email address** if no suitable answer is found.
- **Self-improves** over time by logging interactions and integrating new data (self-learning model).

📚 Dataset fourni

Source : Questions-Export-2025-October-27-1237 (1).xlsx
Format: Excel (.xlsx)
Volume: ~400 rows (questions, answers, topics, metadata)

Important:
Your work may be reused by the PLV IT department, so please keep in mind:

- **Expected output:** response in **HTML format**
- **Priority:** focus on **intelligence and answer relevance**, not on UI design
- **Constraints:** the system must be **reliable, secure**, and **easily integrable** into the student portal (helpcenter)

💡 Bonus – To facilitate integration:

Technologies used by the IT department: **Vue.js** (front-end) and **MariaDB** (database).

🔧 Expected Deliverables (NON-MODIFIABLE)

- Project description + code (GitHub repository)
- Pitch deck (.pdf)
- Demo video (≤ 5 min)

👩‍💼 Expert Contact Point

- **Name** : Guillaume Douceron
- **Position** : Digital Transformation Project Manager
- **Email** : guillaume.douceron@devinci.fr

DEVINCI JUNIOR
Paris - Nantes