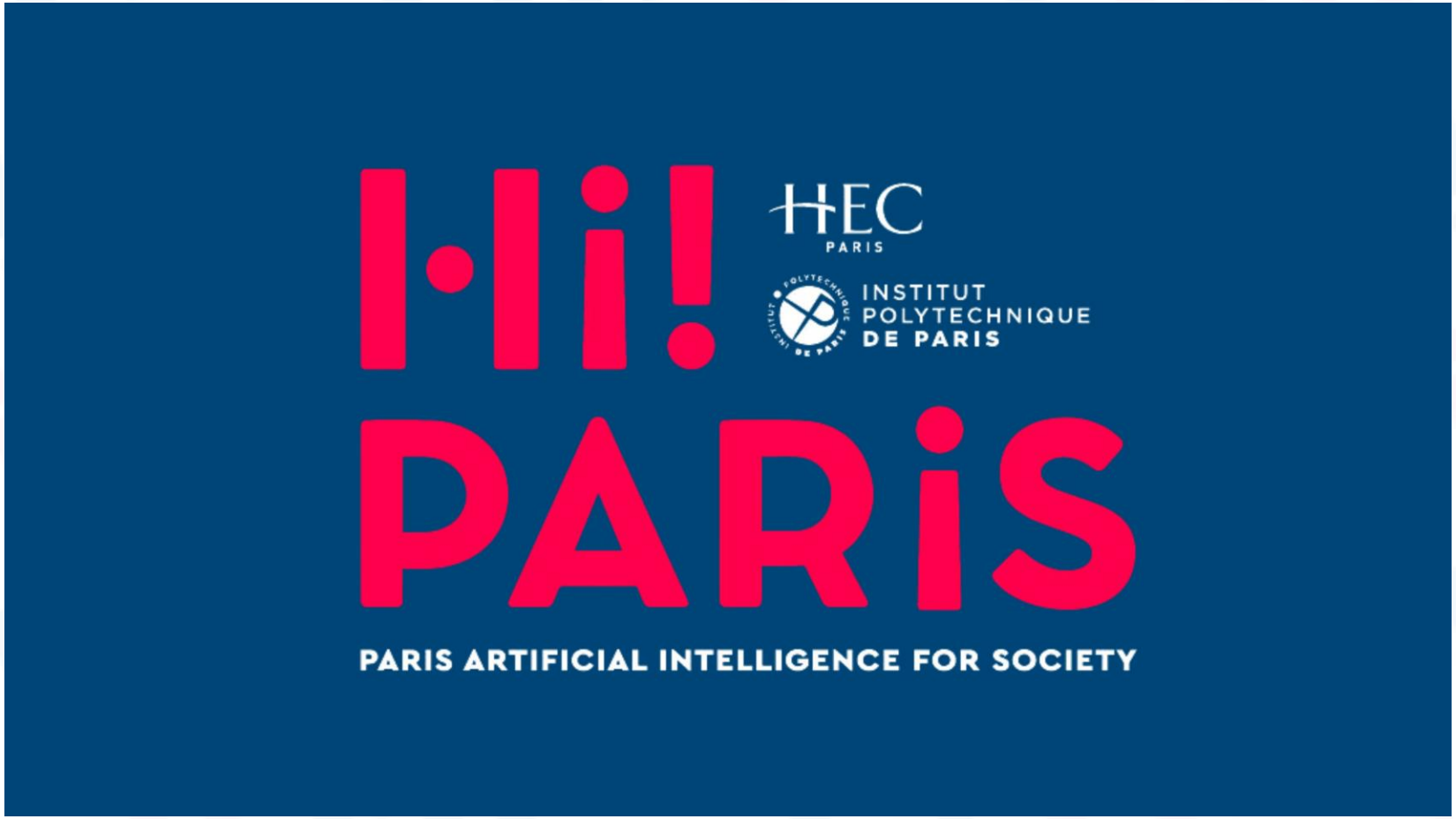


Turbofan Failure Prediction

Team Introduction – CPES Saclay & X

5 students with complementary skills in Data Science, ML, DL & Business.

Raphaël CONTRI

Demartino KUETE TSAYO

Othmane NAMMOUS

Tharushan UTHAYAKUMAR

Rémi MALAPERT

Data Preparation & Exploration

Data Preparation & Exploration

❑ **Dataset:** NASA C-MAPSS FD001 (cleaned in the 1st TP : 3 columns had missing values)

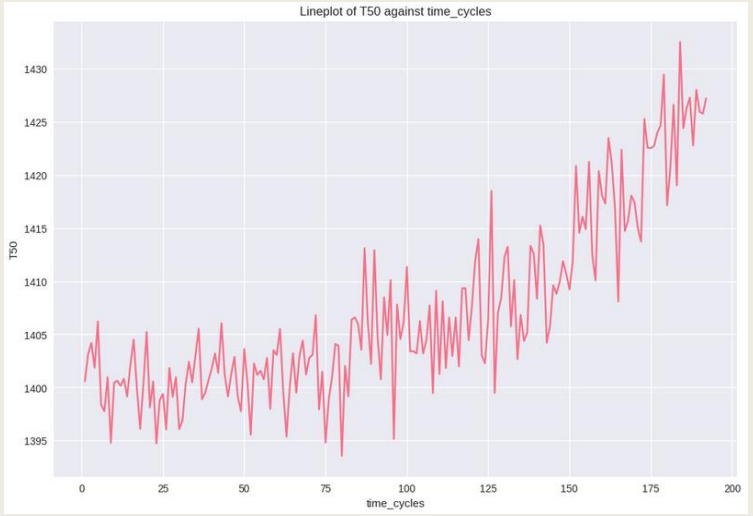
100 training engines, 100 test engines

26 sensor features + operational settings

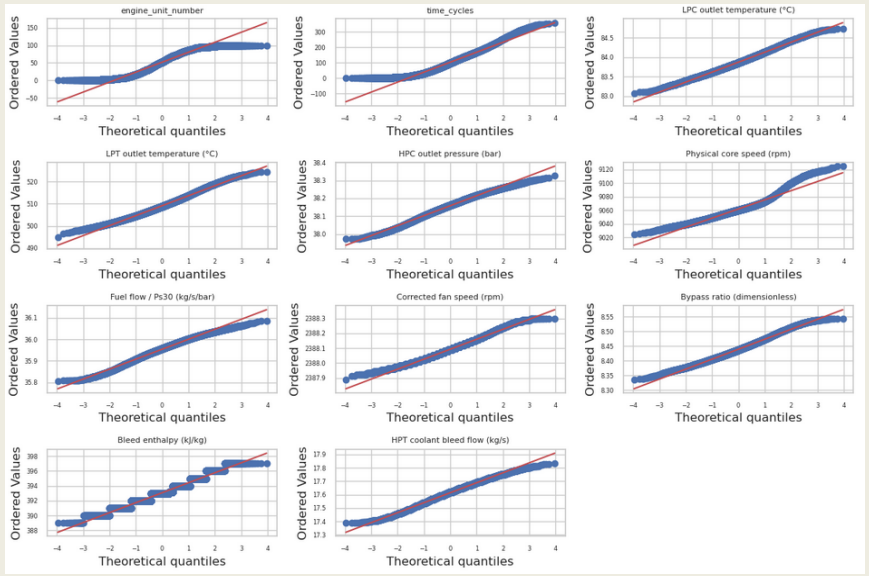
Time-series → target: Remaining Useful Life (RUL)

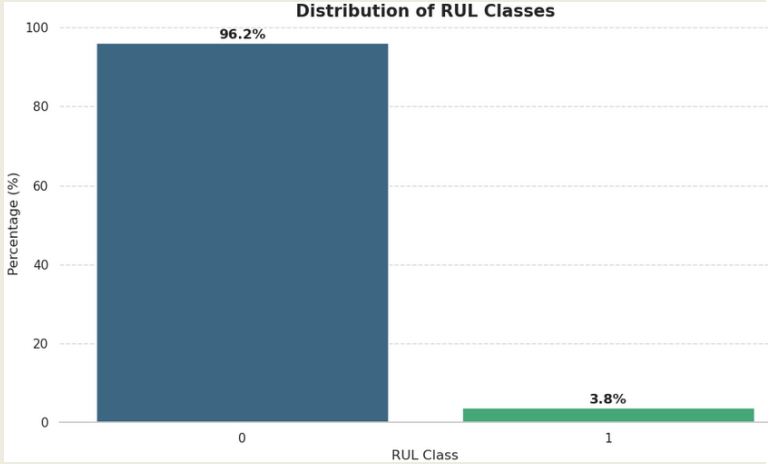
❑ **Challenges:**

Class imbalance (only 3.8% class 1 on the test set).



T50 sensor is positively correlated to time_cycles





❑ **Exploration findings:**

Q-Q plot to check that data follow a normal distribution.

One categorical variable to encode : HPC outlet temperature(°C)

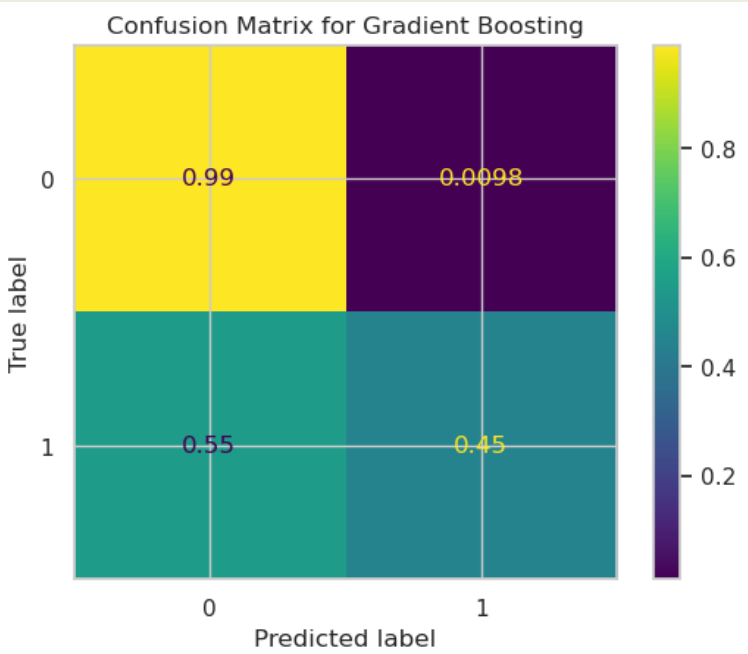
Strong correlation between operational cycles and degradation.

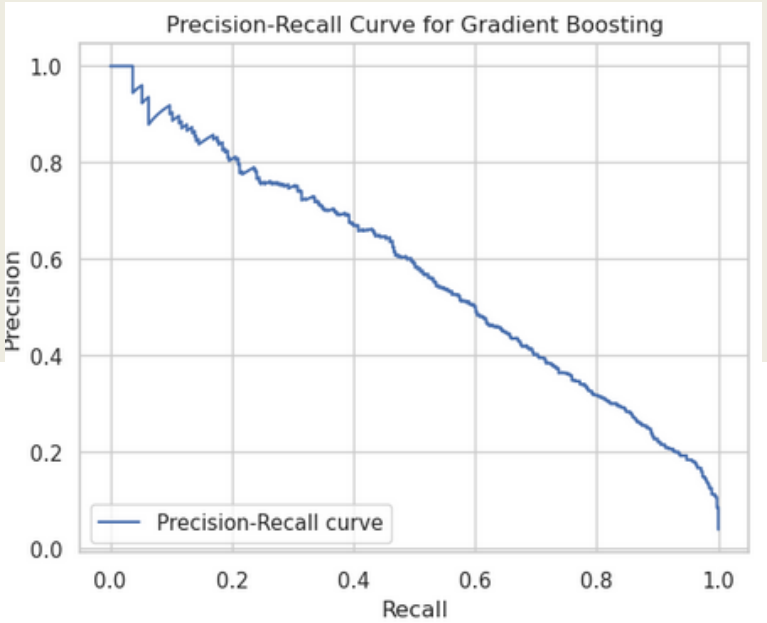
Modeling Approach

Modeling Approach

- **Models tested:** Decision Tree, KNN, Gradient Boosting.
- **Evaluation metric:** Accuracy > 90% for all models → Recall prioritized. It is weak because of imbalance in RUL_class.
- **Best model:** Gradient Boosting (highest recall, most robust).
- **Optimization:** hyperparameter tuning (learning rate, max depth, n_estimators). It was not convincing, so we kept the first gradient boosting model.
- **Precision-Recall Curve**

Model	Test recall
Decision Tree	0.38177874186550
KNN	0.42082429501084
Gradient Boosting	0.44902386117136





Confusion matrix insights :
• Only about 1% of class 0 instances are misclassified as 1 : very few false positives.
• 55% of class 1 (minority class) instances are wrongly classified as 0 : many false negatives.

Precision recall curve insights:

Explainability & Key Drivers

Explainability & Key Drivers

- **SHAP values (top features):**

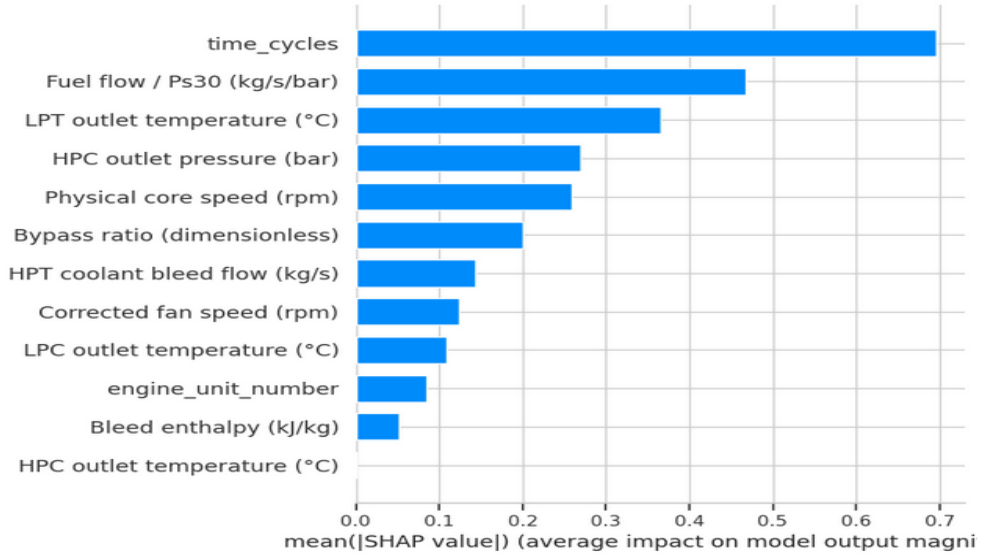
Operational cycles (~0.70).

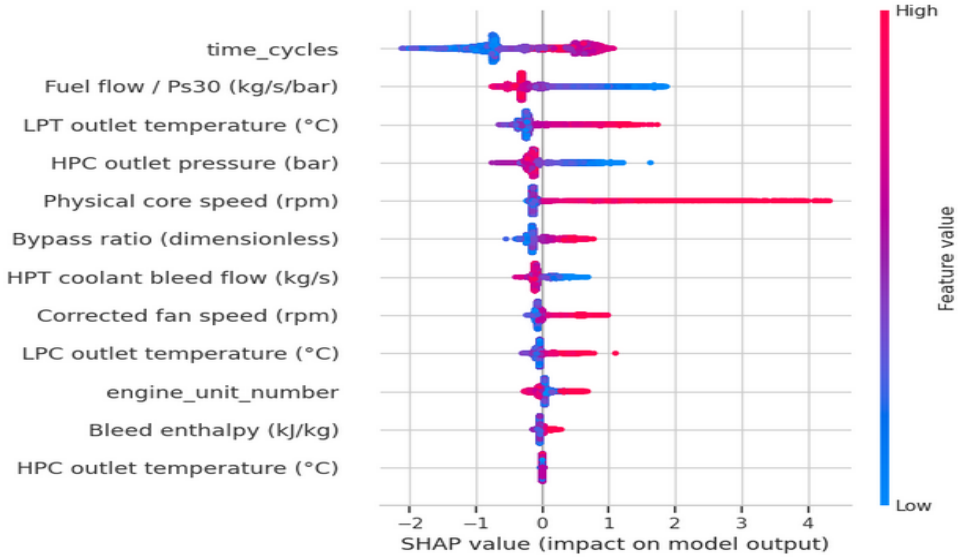
Fuel flow (~0.47).

LPT outlet temperature (°C) (~0.37).

- **Key insight:**

As time passes (the higher the time cycle value is), RUL_class becomes 1 (the engine is not useable). The more the fuel flow is, the closer to 0 the RUL_class is (and it is the opposite for LPT outlet temperature. For LPT outlet temperature, we see the correlation we found (slide 4)).

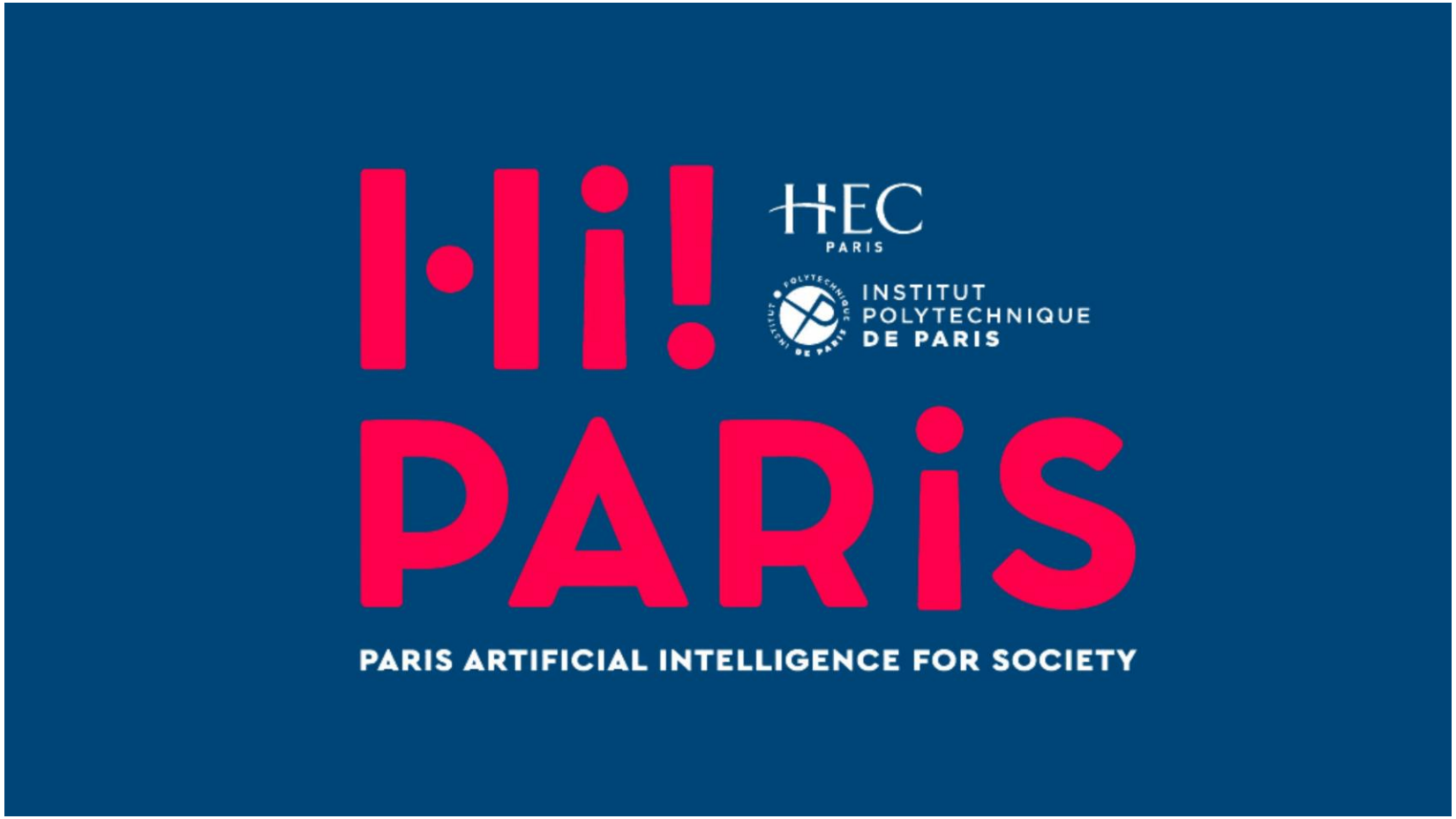




Conclusions & Next Steps

Conclusions & Next Steps

Strenghts	Limitations	Recommendations
Solid preprocessing pipeline.	Strong class imbalance → recall still relatively low.	Apply resampling/SMOTE to address imbalance.
Gradient Boosting achieved the best recall even if it remains low.	The neural network we tried was not convincing : the results were not better than with Gradient Boosting.	Test other classification models like XGBoosti or AdaBoost
Explainability provided interpretable drivers of degradation.		


HEC
PARIS
INSTITUT
POLYTECHNIQUE
DE PARIS
PARIS
PARIS ARTIFICIAL INTELLIGENCE FOR SOCIETY

Thank you !