



RIOS

CHANNEL

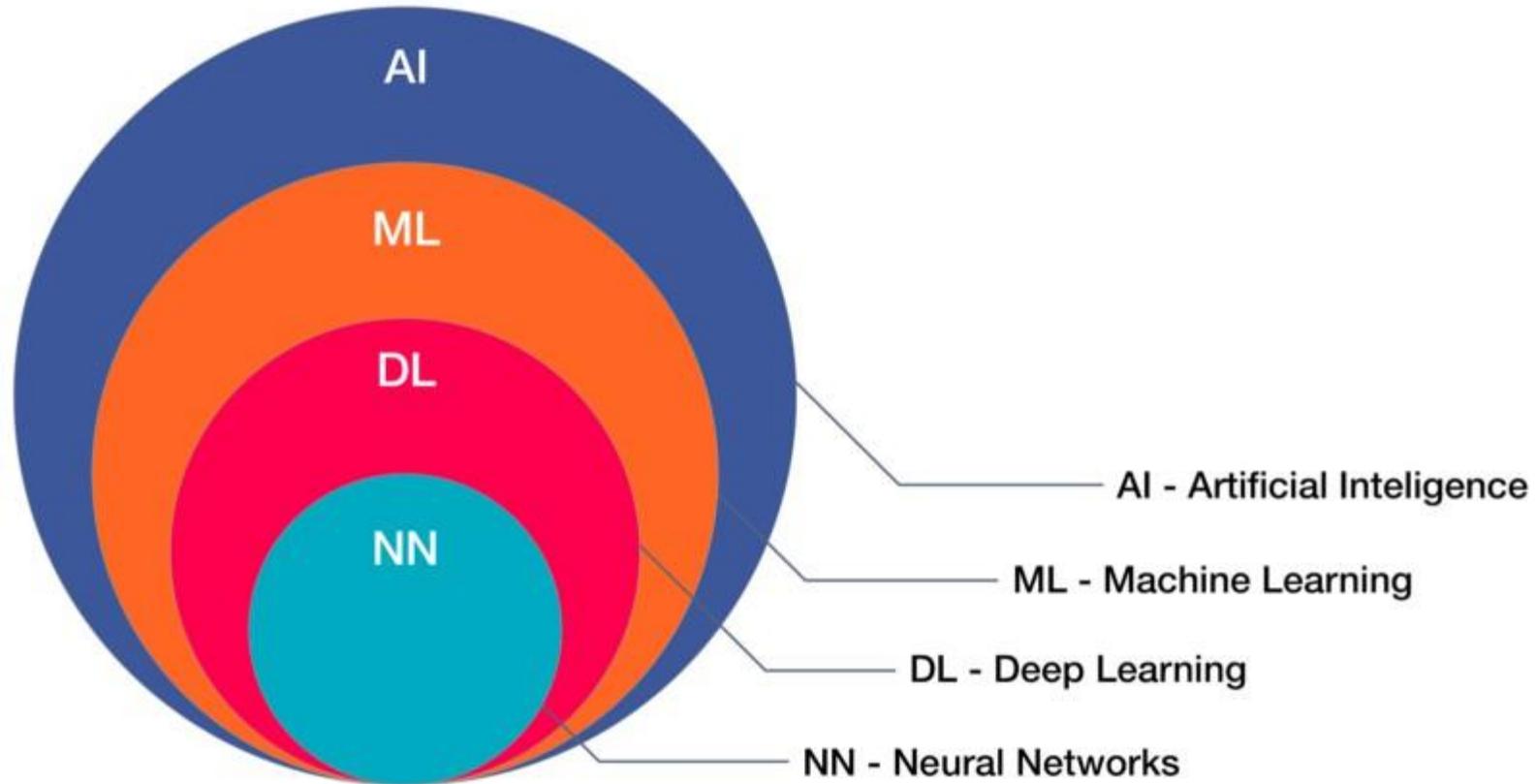
Martedì 7 luglio 2020



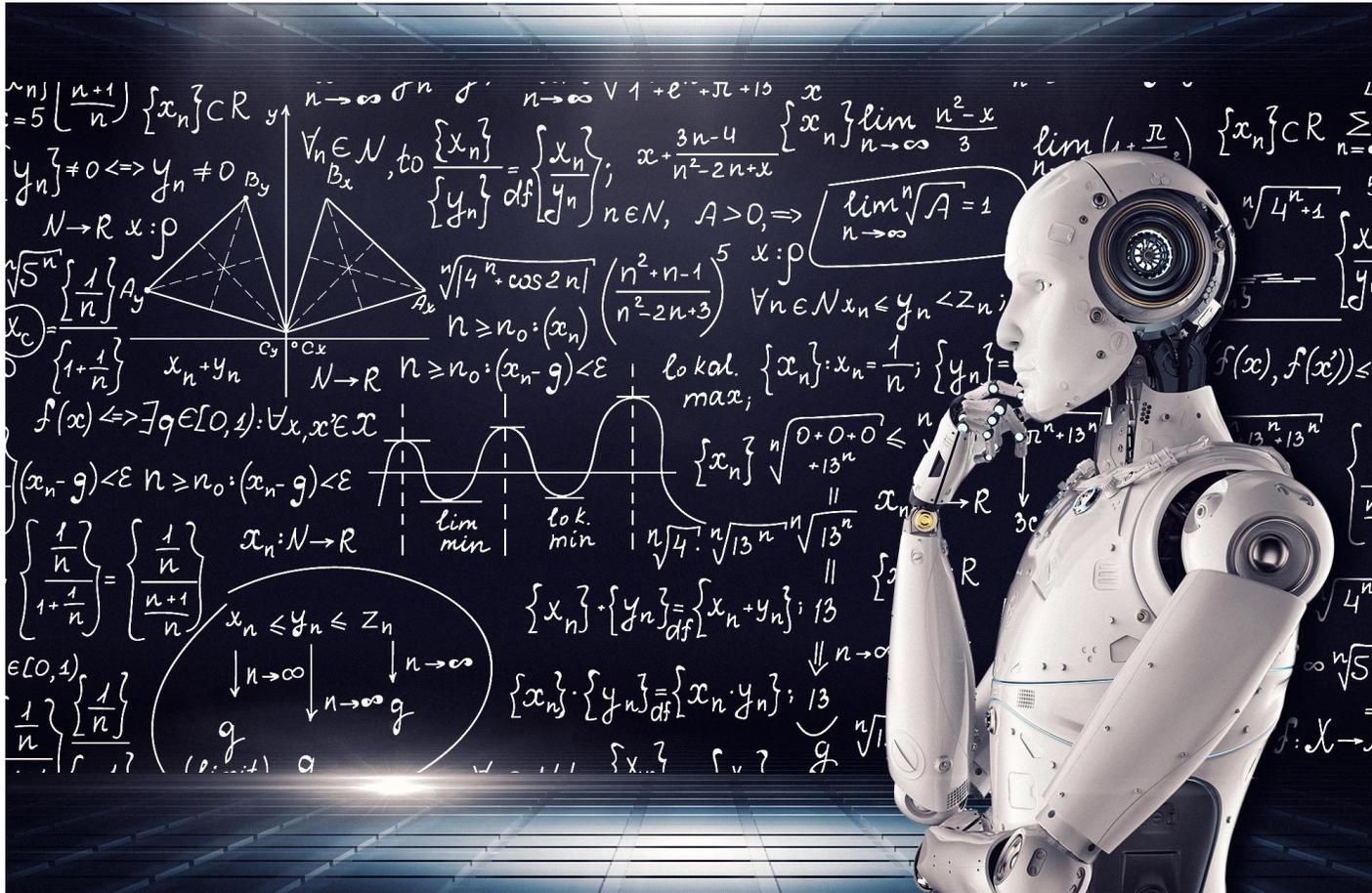
CHANNEL

MACHINE LEARNING E INTELLIGENZA ARTIFICIALE

MACHINE LEARNING E INTELLIGENZA ARTIFICIALE



MACHINE LEARNING



Il **machine learning** può essere considerato una variante alla programmazione tradizionale, nel quale si predispone in una macchina l'abilità di **apprendere dai dati in maniera autonoma**, senza essere esplicitamente programmata per farlo, per poi poter riutilizzare ciò che è stato appreso (**esperienza**) su un certo compito da svolgere.

Questo avviene attraverso l'utilizzo di **modelli matematici** come le reti neurali artificiali, gli algoritmi adattativi, l'elaborazione delle immagini, il data mining, e altri.

APPLICAZIONI DEL MACHINE LEARNING NELL'INDUSTRIA

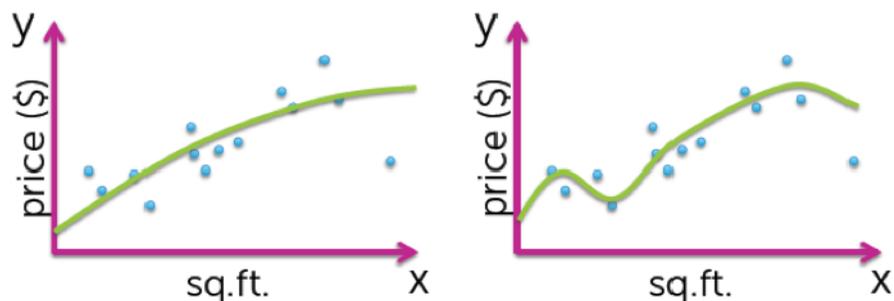


TIPOLOGIE DI MACHINE LEARNING: SUPERVISED LEARNING

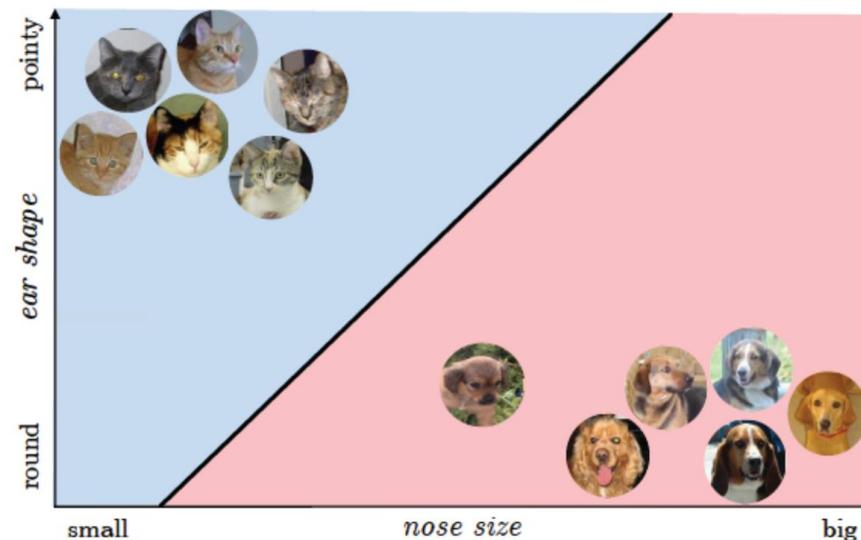
In questa tipologia di machine learning i modelli vengono addestrati a partire da dati etichettati, cioè dati per cui sono conosciuti i segnali di output desiderati.

Ciò che il modello apprende sono una serie di pattern e correlazioni tra il dato e l'etichetta associata, che riutilizzerà come base per prendere decisioni su dati futuri.

REGRESSIONE



CLASSIFICAZIONE

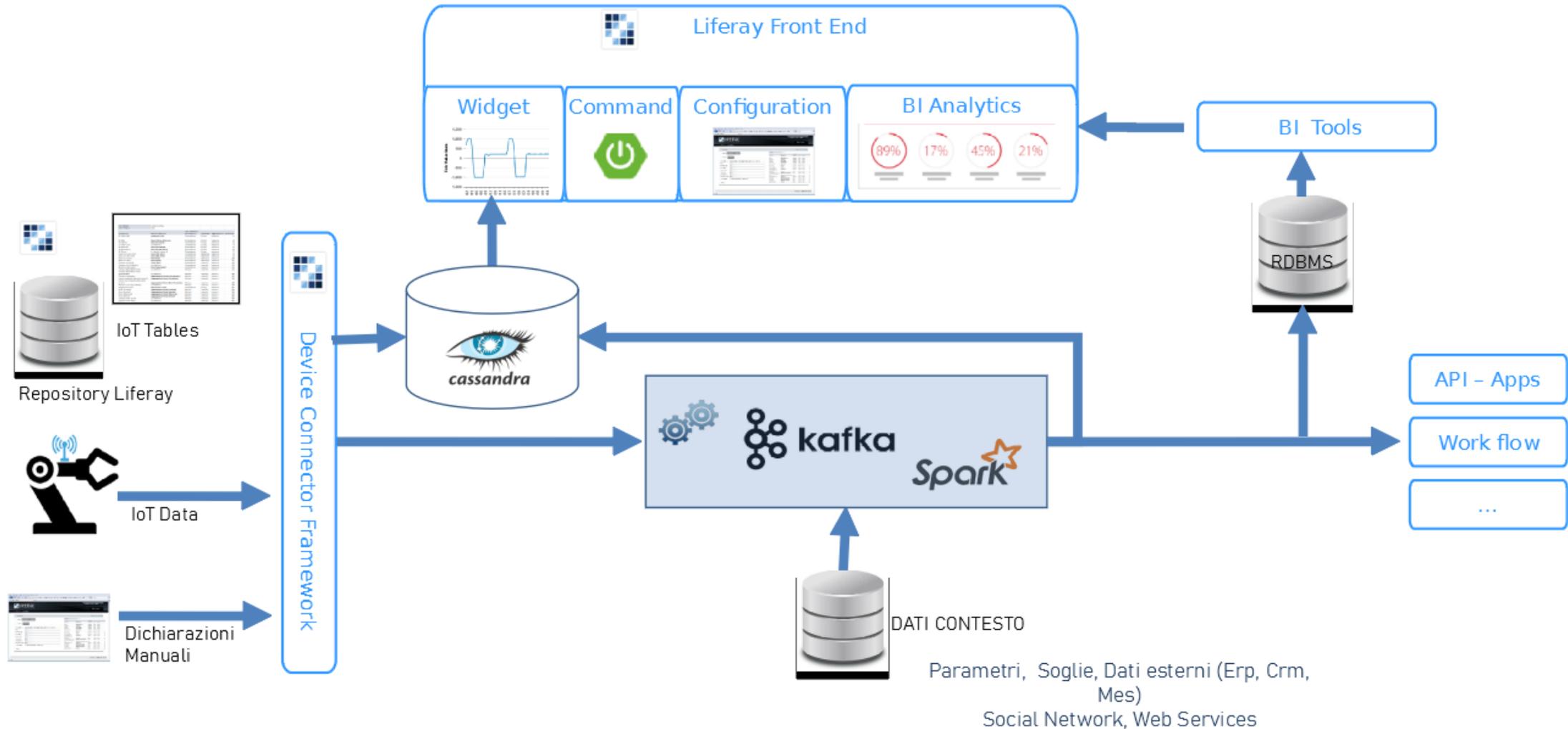




CHANNEL

CASO D'USO: MANUTENZIONE PREDITTIVA CON DATI IOT

ARCHITETTURA - IOT EXPERIENCE



OBIETTIVI PER LA MANUTENZIONE PREDITTIVA

- Previsione di quando potrebbe verificarsi il prossimo guasto (Regressione)
- Previsione se un guasto possa verificarsi in un certo periodo (Classificazione)
- Rilevazione di anomalie nel funzionamento
- Previsione delle variazioni del rischio di un guasto (degradamento del processo)



COSTRUZIONE DEL DATASET

ACQUISIZIONE DEI DATI

- Sono stati acquisiti i dati relativi ai guasti storici che si sono verificati sul macchinario
- in un periodo di 9 mesi

- Inoltre è stato acquisito tutto lo storico della telemetria nel periodo di tempo precedente ai guasti considerati.

FEATURE ENGINEERING

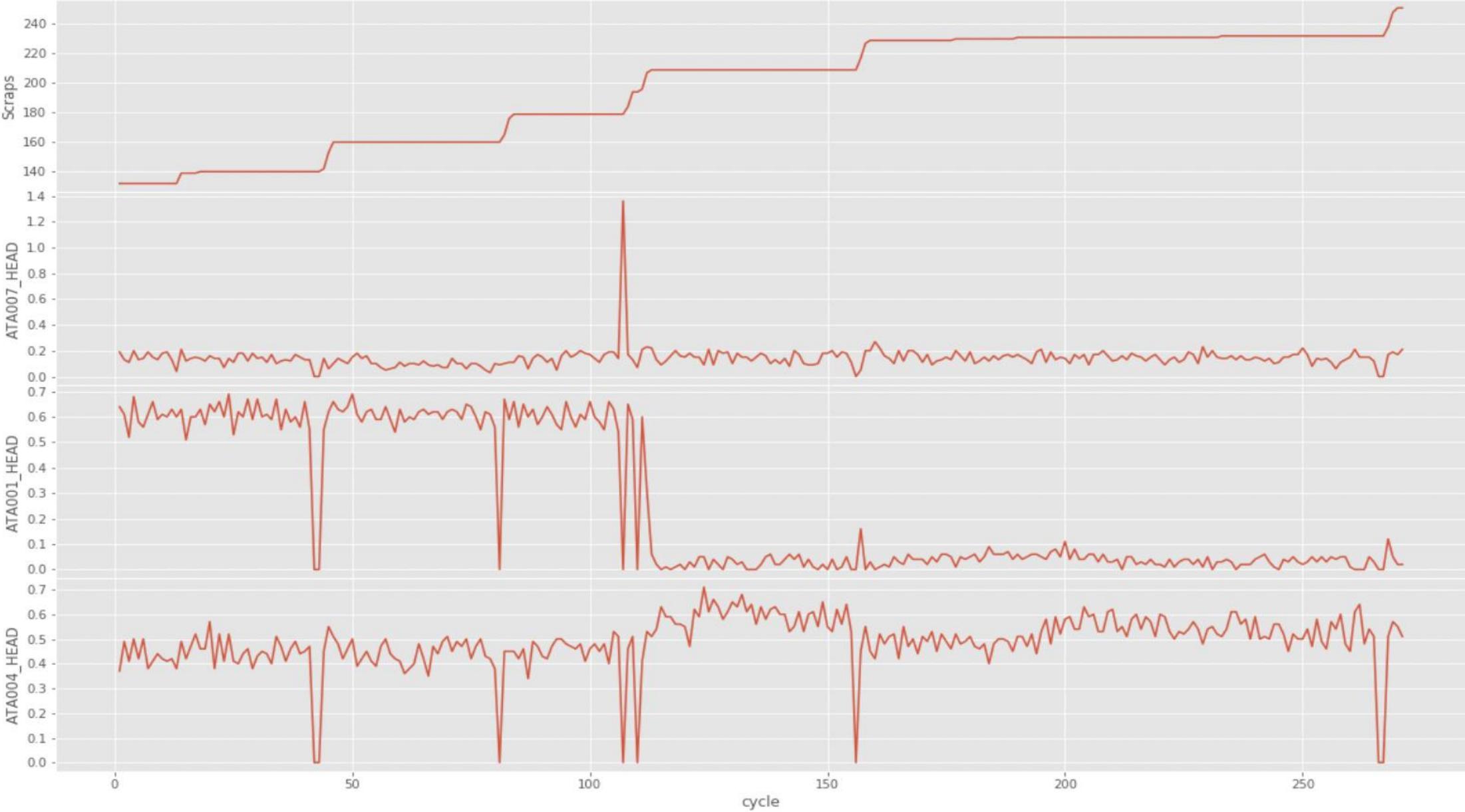
- Filtraggi, elaborazioni e aggregazioni

ETICETTATURA

- I dati sono stati etichettati con due diversi valori per realizzare un approccio di tipo supervisionato:
 - Un **valore su scala continua**, che specifica il tempo rimanente al guasto nel momento in cui è stato registrato nel record telemetrico
 - Un **valore binario**, per il task di classificazione binaria, che indica se il record telemetrico è stato registrato all'interno di una certa finestra temporale dal guasto.



DATI TELEMETRICI



TEST DI ADDESTRAMENTO

I test sono stati svolti utilizzando gli strumenti messi a disposizione in un ambiente IBM Cloud:

- Watson Studio
- Watson Machine Learning

The screenshot displays the IBM Watson Studio interface for an experiment named 'binary_classification_modified_30'. The top navigation bar includes the IBM Watson Studio logo, a refresh button, and the user 'SMC Treviso S.r.L.' with a dropdown menu and a 'DC' indicator. The breadcrumb trail shows 'My projects / Sperimentazioni ML Zoppas / binary_classification_modified_30'. The main content area is divided into two tabs: 'Experiment summary' (selected) and 'Pipeline comparison'. The 'Experiment summary' tab shows a workflow diagram with steps: Read dataset, Split holdout data, Read training data, Preprocessing, Model selection, Extra Trees Classifier, Hyperparameter optimization (P1), Feature engineering (P2), Hyperparameter optimization (P3), and Hyperparameter optimization (P4). A 'Rank by: Accuracy (Optimized)' dropdown and a 'Score: Cross validation' dropdown are visible. A 'Holdout' button is also present. On the right, a circular progress indicator shows the experiment is completed, with 4 pipelines generated. A message states: 'Experiment completed 4 PIPELINES GENERATED 4 pipelines generated from algorithm. See pipeline leaderboard below for more detail. Time elapsed: 6 minutes'. A 'View full log' link is at the bottom right of this section. Below the workflow, the 'Pipeline leaderboard' section contains a table with the following data:

Rank	Name	Algorithm	Accuracy (Optimized)	Enhancements	Build time
★ 1	Pipeline 3	Extra Trees Classifier	0.978	HPO-1 FE	00:02:05
2	Pipeline 4	Extra Trees Classifier	0.978	HPO-1 FE HPO-2	00:01:15
3	Pipeline 1	Extra Trees Classifier	0.970	None	00:00:03
4	Pipeline 2	Extra Trees Classifier	0.970	HPO-1	00:00:37

ESECUZIONE DEI TEST

I dati sono stati suddivisi in due insiemi:

- 90% training (usato nella fase di addestramento)
- 10% test (usato per valutare la performance del modello su dati "nuovi")

Algoritmi sperimentati:

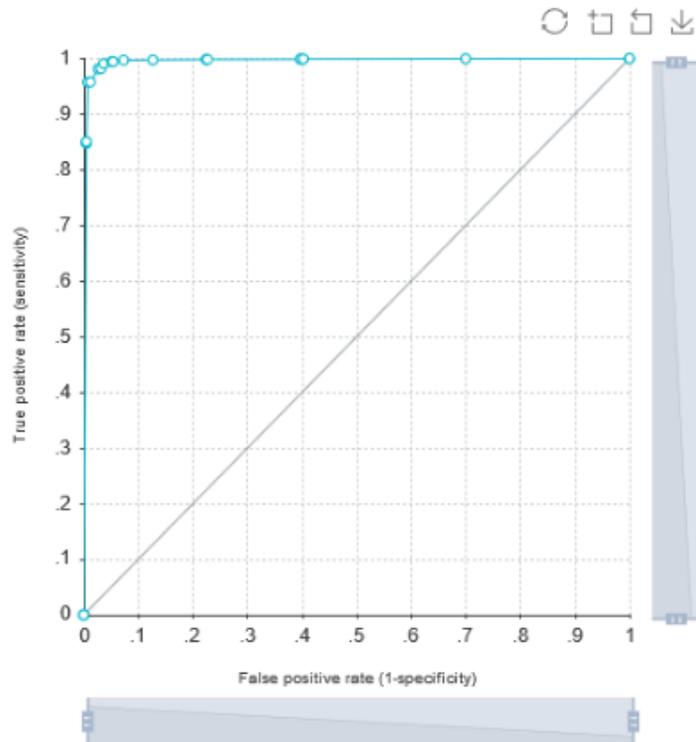
- LGBM Regressor (Light Gradient Boost Machine)
- Extra Trees Regressor
- Random Forest Classifier
- LSTM (reti neurali di tipo Long Short Term Memory)

VALUTAZIONE DEI RISULTATI

Model Accuracy



ROC Curve 



Model Evaluation Measures

	Holdout Score	Cross Validation Score
Accuracy	0.986	0.971
Area Under ROC Curve	0.995	0.991
Precision	0.982	0.963
Recall	0.927	0.840
F ₁ Measure	0.954	0.897
Average Precision	0.979	0.960
Log Loss	0.094	0.121

VALUTAZIONE DEI RISULTATI

Confusion Matrix ⓘ

TARGET : LABEL_BNC

Observed	Predicted		
	0	1	Percent Correct
0	2,283	7	99.7%
1	30	380	92.7%
Percent Correct	98.7%	98.2%	98.6%

Less correct  More correct

- Nel 92,7% dei casi se siamo vicini a un possibile guasto il sistema è in grado di prevederlo correttamente
- il 98,2% delle volte che viene segnalato un possibile imminente guasto questo accadrà realmente, mentre nel 1,8% dei casi si tratterà di un falso allarme
- Complessivamente nel 98,6% dei casi otteniamo una previsione corretta

POSSIBILI IMPLEMENTAZIONI





BREAKFAST CHANNEL

Grazie



Segui @RIOS_opensource



<http://www.reteitalianaopensource.net>