MANAGE AI
From SAFE AI to Credem Internal Framework

Mattia Centurelli

PARALLEL SESSION 3.2
AI e Cyber Risk interrogano il rischio operativo
AI and Cyber Risk – How they bear on operational risk

Framework KPI AI

SAFE ARTIFICIAL INTELLIGENCE

SUSTAINABILITY

It should be **efficient and robust**, in terms of data and computations

ACCURACY

It should lead to **accurate predictions**

FAIRNESS

It should **not discriminate** by age, ethnicity, gender or other population groups

EXPLAINABILITY

It should be **interpretable** in terms of its drivers

Framework KPI AI

ScienceDirect

SAFE ARTIFICIAL INTELLIGENCE

SUSTAINABILITY

It should be **efficient and robust**, in terms of data and computations

ACCURACY

It should lead to **accurate predictions**

FAIRNESS

It should **not discriminate** by age, ethnicity, gender or other population groups

EXPLAINABILITY

It should be **interpretable** in terms of its drivers

Da SAFE al contesto indicatori Credem

INDICE DI IDONEITÀ

Indice che misura l'idoneità del modello di Analytics Governance rispetto **all'attivazione dei ruoli previsti** e del **presidio degli algoritmi censiti**.

- **INDICE DI ATTIVAZIONE:** misura l'effettiva attivazione dei ruoli di Analytics Governance e dei presidi necessari alla governance degli Analytics.
- **INDICE DI GOVERNO:** misura, attraverso opportuna checklist, l'attivazione dei presidi di governance necessari rispetto agli algoritmi censiti.
- **INDICE DI ESERCIZIO:** basato su metriche di performance e fairness relativamente a soglie definite.

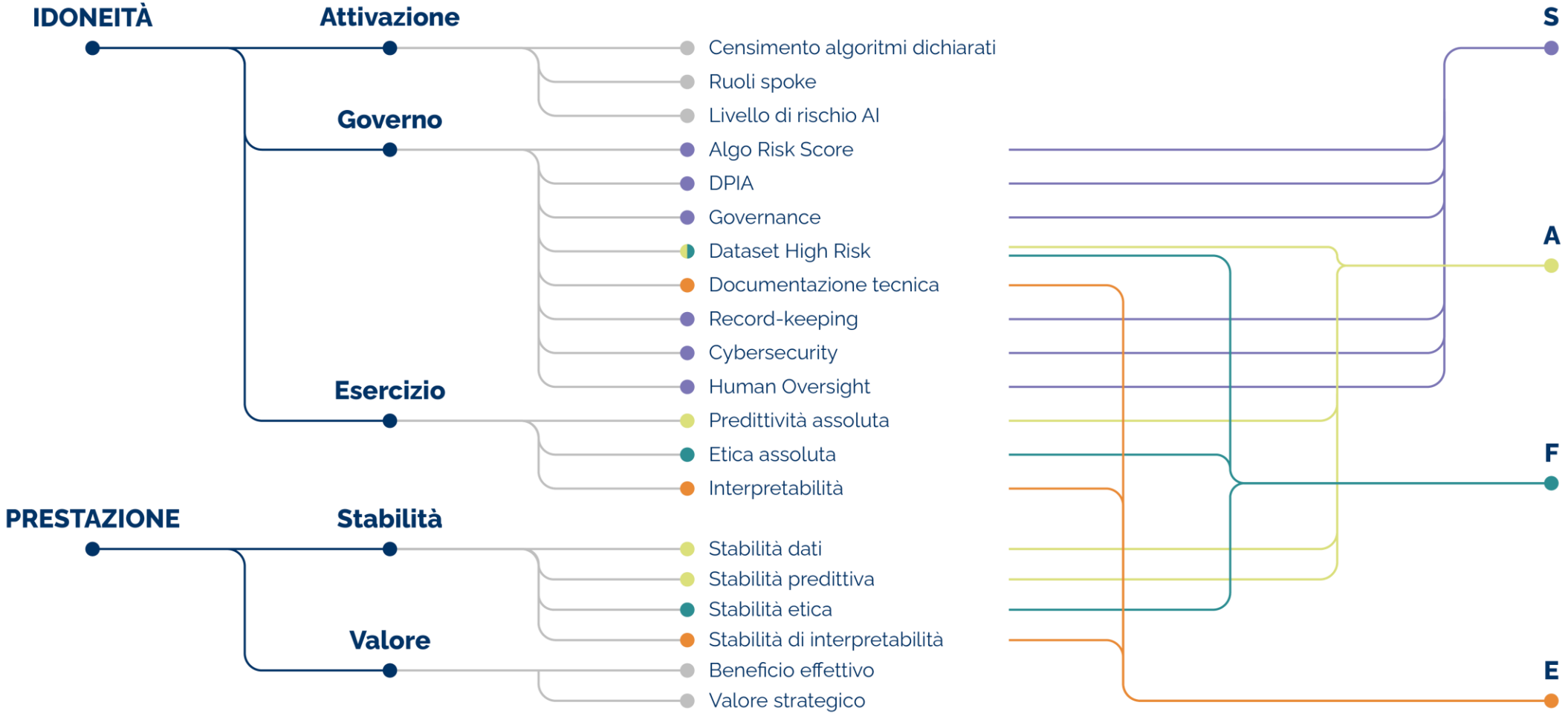
INDICE DI PRESTAZIONE

Indice che monitora la prestazione del modello di Analytics Governance in termini di **performance e valore**.

- **INDICE DI STABILITÀ:** misura la stabilità nel tempo delle metriche di performance, fairness ed explainability.
- **INDICE DI VALORE:** misura il valore effettivamente generato dall'implementazione di soluzioni di Advanced Analytics rispetto alle aspettative.

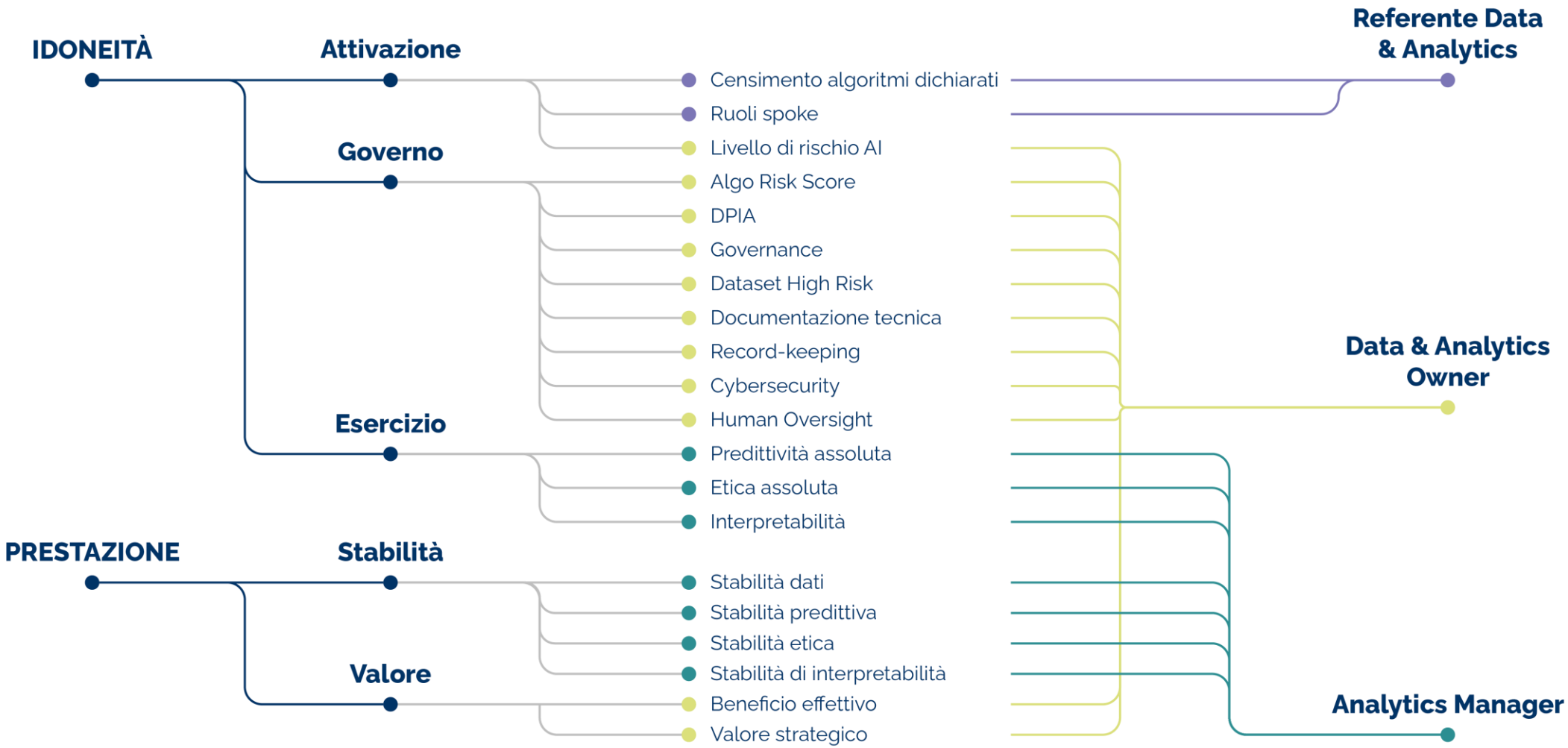
Struttura indicatori macro

Il presente monitoraggio **si applica a tutti gli algoritmi censiti sul modello** di Data & Analytics Governance.



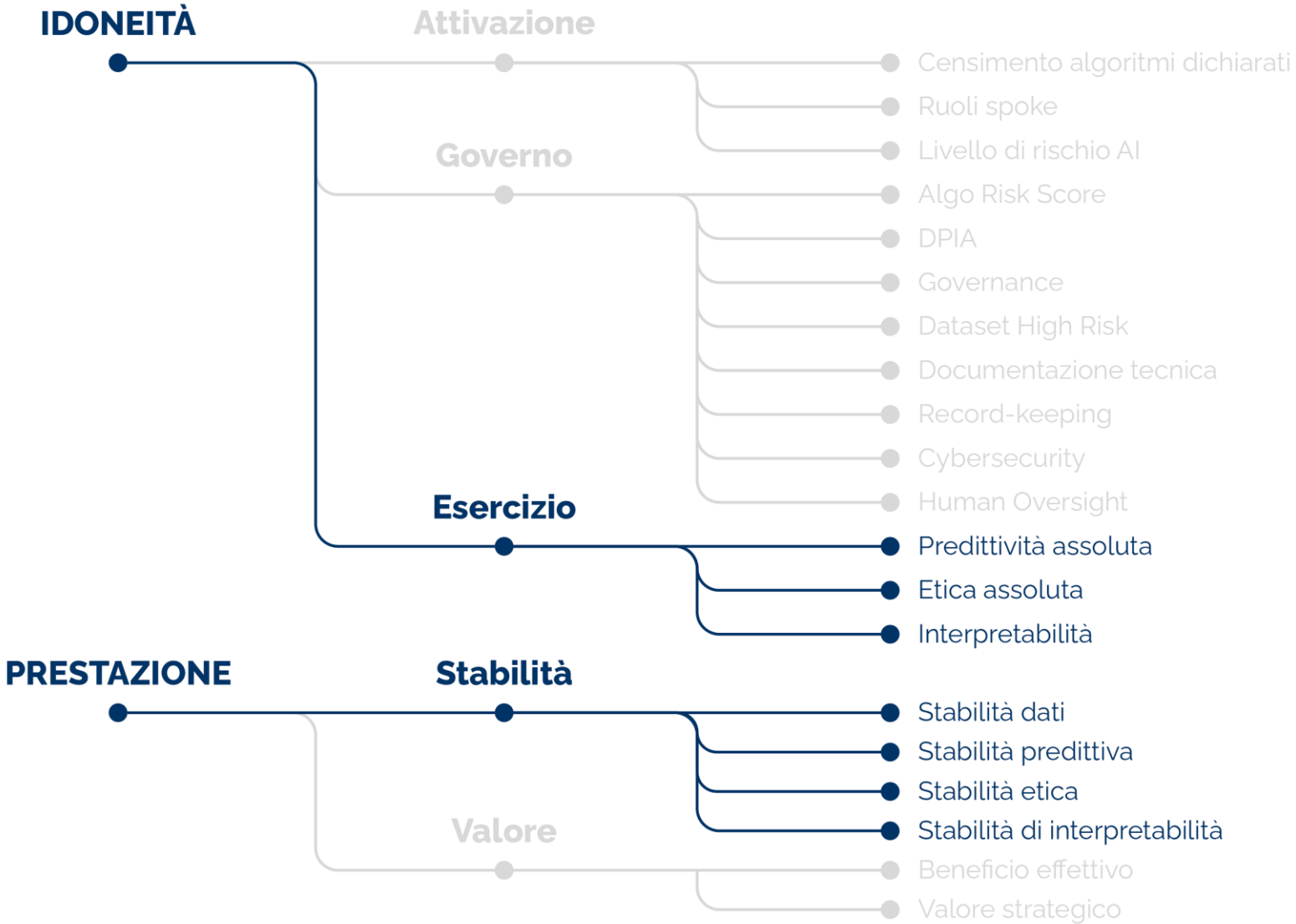
Struttura indicatori macro

Il presente monitoraggio **si applica a tutti gli algoritmi censiti sul modello** di Data & Analytics Governance.



MODEL MONITORING

Indicatori implementati



Model Monitoring - Struttura

Il framework per implementare il **monitoring dei sistemi di ML** è composto da **8 moduli**, fornendo come output report dettagliati in termini di Performance, Data-Drift Fairness e XAI, e un report sintetico contenente gli esiti degli indicatori di Data & Analytics.

PerformanceMeasures: computa le performance di un modello nelle metriche fornite in input.

FairnessMeasures: computa le performance di un modello nelle metriche di fairness fornite in input.

PerformanceDrift: computa il drift delle performance tra dati correnti e dati storici e genera un report con un sistema di alerting.

FairnessDrift: computa il drift delle performance in termini di Fairness tra dati correnti e dati storici e genera un report con un sistema di alerting,

ReferenceMetadata: computa un dizionario di metadati con le caratteristiche statistiche fondamentali a partire da un set di dati o da uno di riferimento.

XAI: computa la spiegabilità di un modello tramite lo scoring delle feature nelle predizioni.

DataDrift: computa il drift dei dati tramite PSI (Population Stability Index) tra dati correnti e dati storici e genera un report con un sistema di alerting.

XAIDrift: computa il drift nella spiegabilità del modello tra dati correnti e dati storici e genera un sistema di alerting.

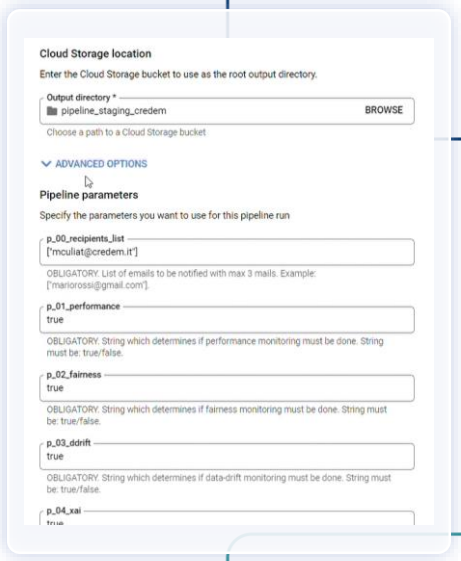
Model Monitoring – Tools

Il framework di monitoraggio viene implementato tramite una **Vertex AI Pipeline parametrizzabile**.

Cloud Big Query: Data Warehouse

Cloud Artifact Registry: Docker & Kubeflow Pipeline Repositories

Cloud Storage: Storage Metadati e Risorse delle Pipeline



Input

Cloud Vertex AI: Vertex AI Pipelines

Output

Cloud Big Query: Report ed Esiti Indicatori

Cloud Storage: Storage Metadati e Risorse delle Pipeline

CASO D'USO

Use case: Consumi energetici

ENERGY CONSUMPTION FORECASTING AND ANOMALY DETECTION

Un business case per **predire i consumi energetici e i costi relativi** di tutte le strutture Credem strumentali, **analizzare i consumi anomali e valutare i risparmi**.

Bisogno

Sulla base dello storico dei consumi energetici, predire i costi in fase di budget (annuale) e in fase di consuntivazione (mensile), con simulazione dei costi a mercato, clustering in base ai consumi e anomaly detection.

Potenziali risultati

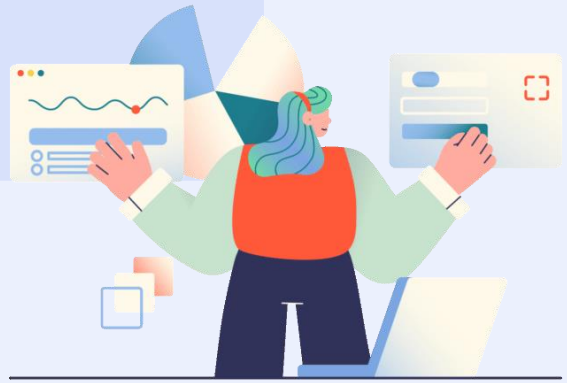
Forecasting dei consumi, simulazione dei costi a mercato, Identificazione di costi recuperabili nel breve e lungo termine.

Per chi

Funzione LOGIS

Strumenti

Google Cloud Platform
Denodo
PowerBI



Solution: Governance

Situazione attuale

6 job schedulati giornalmente attivati ogni volta che ci sono nuovi dati (Data Quality, Data Preparation, Anomaly Detection, Monitoring Pre-Prediction, Modeling, Monitoring Post-Prediction).

Sistema di logging
(Art. 12 per High Risk AI Systems)

Generazione dei logs di ogni singolo job a ogni run.

Solution: Governance

Situazione attuale

6 job schedulati giornalmente attivati ogni volta che ci sono nuovi dati (Data Quality, Data Preparation, Anomaly Detection, Monitoring Pre-Prediction, Modeling, Monitoring Post-Prediction).

Sistema di logging
(Art. 12 per High Risk AI Systems)

Generazione dei logs di ogni singolo job a ogni run.

Data Quality
(Art. 10 per High Risk AI Systems)

- **Generazione di un report di data- quality** sintetico sui dati anagrafici.
- **Generazione di alert** relativi a errori di formato e/o valori mancanti per ogni campo.

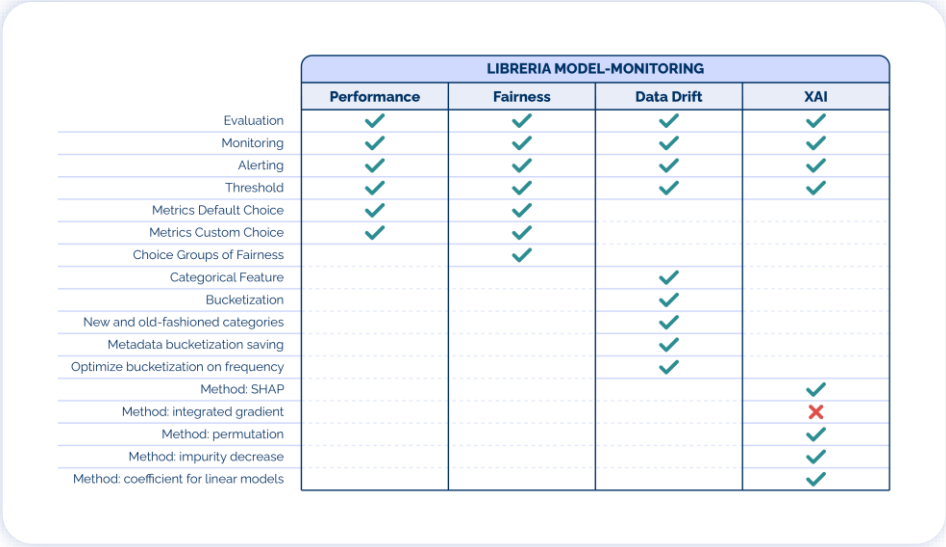
	A	B	C	D
1		Missing	Wrong	Total
2	POD	12	0	12
3	CODICE EDIFICIO	20	3	23
4	TIPO SWITCH 0-1-2	0	0	0
5	RAGIONE SOCIALE	0	0	0
6	CDC PREVALENTE	1	0	1
7	PROVINCIA	2	134	136
8	REGIONE	81	92	173
9	STATO UTENZA	0	0	0
10	SUP. PUL.	0	0	0
11	TIPO IMPIANTO	0	0	0
12	INIZIO VALIDITA'	0	0	0
13	FINE VALIDITA'	0	0	0
14	GIORNI	0	0	0

```
202                                    T ARANCIONE! Valore mancante del CODICE EDIFICIO per il POD IT001E54793289 nella riga 1092. Deve essere: AAXXXXXX
202                                    T ARANCIONE! Valore mancante del CODICE EDIFICIO per il POD IT001E70319794 nella riga 1135. Deve essere: AAXXXXXX
202                                    T ARANCIONE! Valore mancante del CODICE EDIFICIO per il POD IT001E70319794 nella riga 1136. Deve essere: AAXXXXXX
2024-05-09 06:02:27,150 log2       : WARNING  ALERT ROSSO! Formato errato del CODICE EDIFICIO MOXXX1X1 per il POD IT011E10037248 nella riga 1286. Deve essere: AAXXXXXX
2024-05-09 06:02:27,152 log2       : WARNING  ALERT ROSSO! Formato errato del CODICE EDIFICIO MOXXX1X1 per il POD IT011E10037248 nella riga 1287. Deve essere: AAXXXXXX
2024-05-09 06:02:27,153 log2       : WARNING  ALERT ARANCIONE! Valore mancante del CODICE EDIFICIO per il POD IT020E00301027 nella riga 1383. Deve essere: AAXXXXXX
2024-05-09 06:02:27,155 log2       : WARNING  ALERT ARANCIONE! Valore mancante del CODICE EDIFICIO per il POD IT020E00301027 nella riga 1384. Deve essere: AAXXXXXX
2024-05-09 06:02:27,157 log2       : WARNING  ALERT ROSSO! Formato errato del CODICE EDIFICIO VI00101 per il POD IT025E00956538 nella riga 1426. Deve essere: AAXXXXXX
2024-05-09 06:02:27,158 log2       : WARNING  ALERT ARANCIONE! Valore mancante del CODICE EDIFICIO per il POD nan nella riga 1456. Deve essere: AAXXXXXX
2024-05-09 06:02:27,159 log2       : WARNING  ALERT ARANCIONE! Valore mancante del CODICE EDIFICIO per il POD nan nella riga 1461. Deve essere: AAXXXXXX
```

Solution: Governance

Situazione attuale

6 job schedulati giornalmente attivati ogni volta che ci sono nuovi dati (Data Quality, Data Preparation, Anomaly Detection, Monitoring Pre-Prediction, Modeling, Monitoring Post-Prediction).



	LIBRERIA MODEL-MONITORING			
	Performance	Fairness	Data Drift	XAI
Evaluation	✓	✓	✓	✓
Monitoring	✓	✓	✓	✓
Alerting	✓	✓	✓	✓
Threshold	✓	✓	✓	✓
Metrics Default Choice	✓	✓		
Metrics Custom Choice	✓	✓		
Choice Groups of Fairness		✓		
Categorical Feature			✓	
Bucketization			✓	
New and old-fashioned categories			✓	
Metadata bucketization saving			✓	
Optimize bucketization on frequency			✓	
Method: SHAP				✓
Method: integrated gradient				✗
Method: permutation				✓
Method: impurity decrease				✓
Method: coefficient for linear models				✓

Sistema di logging
(Art. 12 per High Risk AI Systems)

Generazione dei logs di ogni singolo job a ogni run.

Data Quality
(Art. 10 per High Risk AI Systems)

- **Generazione di un report di data- quality** sintetico sui dati anagrafici.
- **Generazione di alert** relativi a errori di formato e/o valori mancanti per ogni campo.

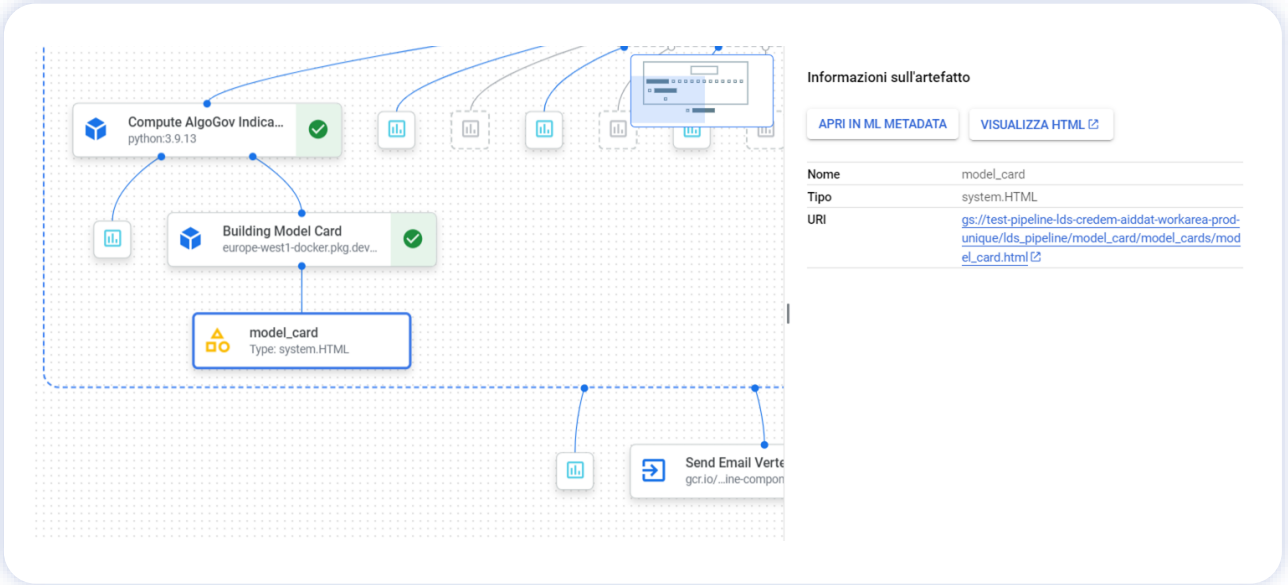
Monitoring

- **Pre-Prediction: Performance e Fairness Monitoring** con i nuovi dati.
- **Post-Prediction: Data-Drift** (con anche le nuove predizioni) e **XAI** dell'output del modello.

Solution: Monitoring

Vertex AI Pipelines

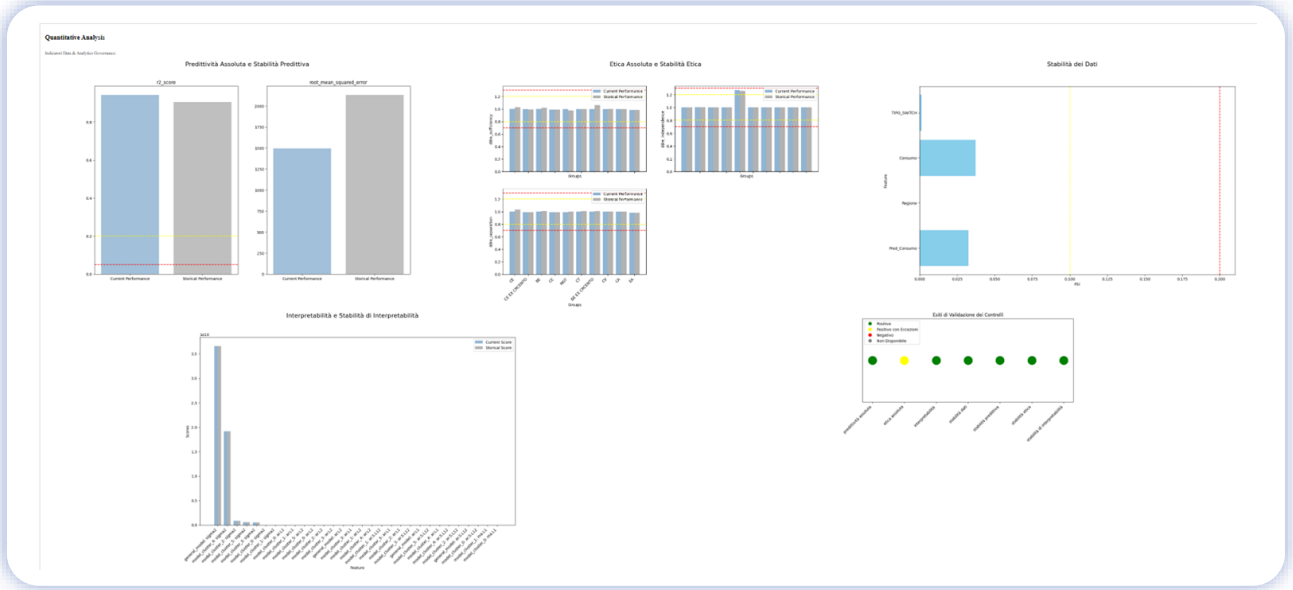
- Aggiornamento o creazione dei baseline metadata su Cloud Storage.
- Generazione dei report di Performance, Fairness, XAI e Data-Drift su Big Query.
- Generazione degli esiti degli indicatori su Big Query.



Solution: Monitoring

Vertex AI Pipelines

- Aggiornamento o creazione dei baseline metadata su Cloud Storage.
- Generazione dei report di Performance, Fairness, XAI e Data-Drift su Big Query.
- Generazione degli esiti degli indicatori su Big Query.
- Generazione di una pagina HTML con una descrizione in generale del monitoring.



Solution: Monitoring

Vertex AI Pipelines

- Aggiornamento o creazione dei baseline metadata su Cloud Storage.
- Generazione dei report di Performance, Fairness, XAI e Data-Drift su Big Query.
- Generazione degli esiti degli indicatori su Big Query.
- Generazione di una pagina HTML con una descrizione in generale del monitoring.
- Storico del monitoraggio.