

Sistemi Autonomi

Open source AutoML: machine learning for everyone?

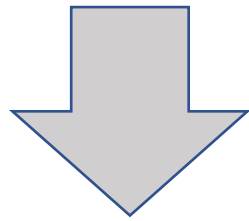
Andrea Lanari

Prof Andrea Omicini, Tutor Didattico Giovanni Ciatto

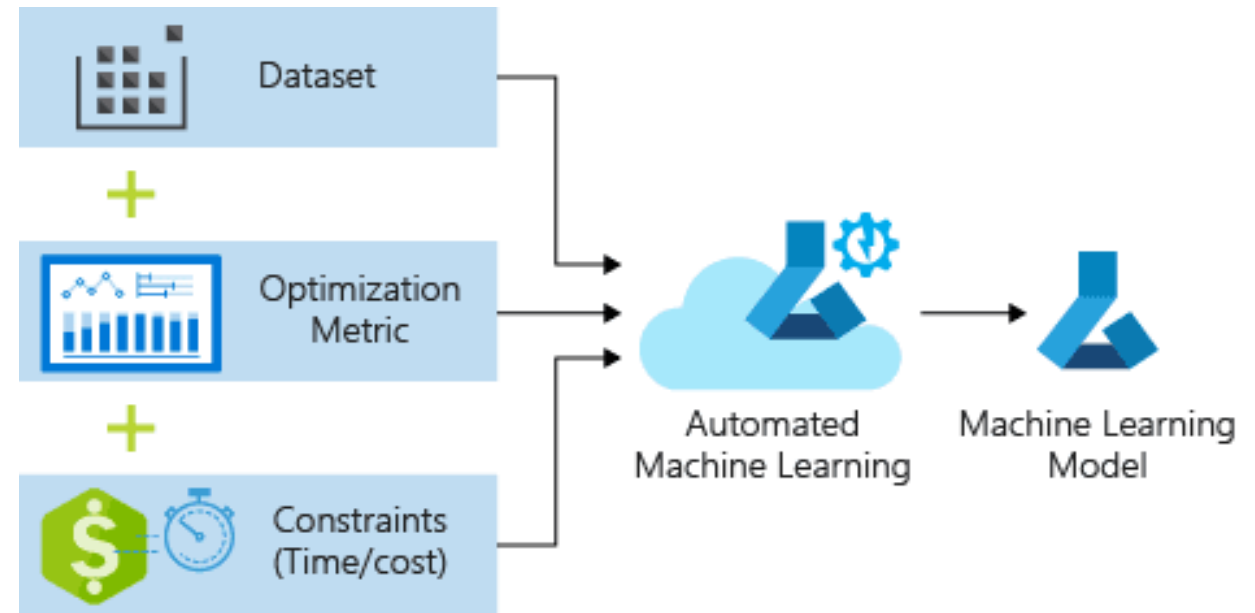


Automated Machine Learning: che cos'è?

Autonomia + Machine Learning



Sistema in grado di compiere
AUTONOMAMENTE uno o più fasi del
processo di *MACHINE LEARNING*



Step del lavoro svolto

1. Si è iniziato facendo una panoramica generale sull'attuale stato dell'arte.
2. Per poi classificare gli strumenti open source sia dal punto di vista dei sistemi autonomi che da quella del machine learning
3. In seguito, si è testato sotto entrambi i punti di vista alcuni dei framework disponibili andando ad analizzare i risultati ottenuti.
4. Infine si è provato a dare una risposta alla domanda presente nel titolo

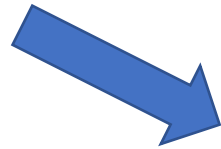
Vari strumenti utilizzano varie metodologie per rendere I lori progetti autonomi:

- *Biology Evolution*
- *Stochastic Behaviours*
- *Greedy Search*
- *Meta Learning\Transfer Learning*

Biology Evolution

- Basati sull'utilizzo di algoritmi genetici
- Esempio: Tpot

*100 alberi casuali
contenuti ML
pipeline
(1° generazione)*



*Selezione dei 20
migliori risultati*



*Creazione di altri
100 elementi
modificando i
selezionati
(n° generazione)*



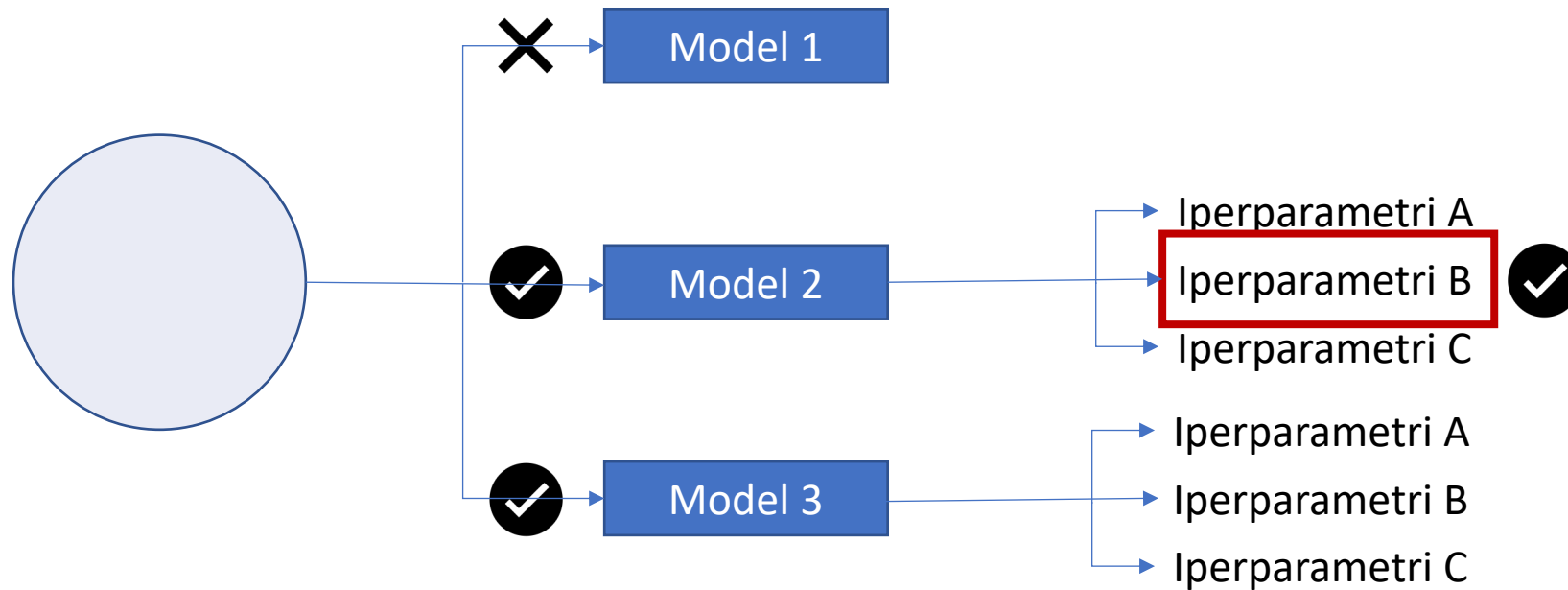
Stochastic behaviours

- Basati sulla costruzione di un modello probabilistico (es. Gaussian Process)
- Ottimizzato attraverso Bayesian Optimization
- Esempio: AutoKeras



Greedy Search

- Ricerca della miglior configurazione mediante la scansione di ottimi locali.
- Esempio: MLJarSupervised



Meta e Transfarn Learning

- Prevede l'uso di Dataset precedentemente analizzati per risolvere il problema
- Esempio: AutoSkLearn (l'ottimizzazione avviene comunque attraverso la Bayesian Optimization)

AutoML e... Machine Learning

- L'obiettivo dell'AutoML è creare strumenti in grado di sostituire l'immagine dell'esperto nel processo di costruzione di un modello di machine learning.
- Si è provato a classificare i vari framework in base alle fasi del processo di Machine Learning che sono in grado di svolgere.
- Considerate 5 categorie.

Classificazione degli Framework

Classe	Descrizione	Esempi
Feature Engineering	Prima fase della Pipeline di machine learning che prevede l'applicazione di diverse tecniche al fine di aumentare le feature che descrivono al meglio il problema.	Explore Kit, FeatureTools

Classificazione degli Framework

Classe	Descrizione	Esempi
Feature Engineering	Prima fase della Pipeline di machine learning che prevede l'applicazione di diverse tecniche al fine di aumentare le feature che descrivono al meglio il problema.	Explore Kit, FeatureTools
Ottimizzazione Iperparametri	Ricerca dei migliori parametri utilizzati dal modello di machine o deep learning (iperparametri)	AutoWeka hyperopt-sklearn

Classificazione degli Framework

Classe	Descrizione	Esempi
Feature Engineering	Prima fase della Pipeline di machine learning che prevede l'applicazione di diverse tecniche al fine di aumentare le feature che descrivono al meglio il problema.	Explore Kit, FeatureTools
Ottimizzazione Iperparametri	Ricerca dei migliori parametri utilizzati dal modello di machine o deep learning (iperparametri)	AutoWeka hyperopt-sklearn
Selezione del modello	Ricerca del miglior modello di machine learning o di (deep learning) in grado di risolvere il sistema.	AutoML_gs ATM (auto tune mode)

Classificazione degli Framework

Classe	Descrizione	Esempi
Feature Engineering	Prima fase della Pipeline di machine learning che prevede l'applicazione di diverse tecniche al fine di aumentare le feature che descrivono al meglio il problema.	Explore Kit, FeatureTools
Ottimizzazione Iperparametri	Ricerca dei migliori parametri utilizzati dal modello di machine o deep learning (iperparametri)	AutoWeka hyperopt-sklearn
Selezione del modello	Ricerca del miglior modello di machine learning o di (deep learning) in grado di risolvere il sistema.	AutoML_gs ATM (auto tune mode)
Neural Architecture Search (NAS)	Feature Engineering + Costruzione della miglior rete neurale che possa risolvere il problema	AutoKeras AutoGluon

Classificazione degli Framework

Classe	Descrizione	Esempi
Feature Engineering	Prima fase della Pipeline di machine learning che prevede l'applicazione di diverse tecniche al fine di aumentare le feature che descrivono al meglio il problema.	Explore Kit, FeatureTools
Ottimizzazione Iperparametri	Ricerca dei migliori parametri utilizzati dal modello di machine o deep learning (iperparametri)	AutoWeka hyperopt-sklearn
Selezione del modello	Ricerca del miglior modello di machine learning o di (deep learning) in grado di risolvere il sistema.	AutoML_gs ATM (auto tune mode)
Neural Architecture Search (NAS)	Feature Engineering + Costruzione della miglior rete neurale che possa risolvere il problema	AutoKeras AutoGluon
General Full Scope	Intero processo di costruzione di un modello di machine learning (Feature Engineering, selezione modello e ottimizzazione iperparametri)	AutoSkLearn MlJarSupervised

L'esperimento proposto

E' un confronto tra possibili tool open source disponibili:

- Solo General Full Scope e i NAS
- 2 Scenari: un problema di classificazione e uno di regressione.
- Valutazioni sia dal punto di vista del ML che dell'autonomia.

I framework presi in considerazione e le loro caratteristiche

Nome Tool	Sviluppato presso	ML classificazione	Strategie per l'autonomia
AutoGluon	Amazon	NAS	Greedy Search
AutoKeras	Keras Team (licenza MIT)	NAS	Bayesian Optimization
AutoSkLearn	Free Project	General Full Scope + Meta Learning	Bayesian Optimization
Ludwig	Uber	NAS	Uso di modelli di default
MLJarSupervised	MLJar (licenza MIT)	General Full Scope	Greedy Search
Tpot	Computational Genetics Lab (University of Pennsylvania)	General Full Scope	Algoritmo Genetico
H2O	H2O.ai	General Full Scope	Greedy Search

Le metriche di ML utilizzate

- Mean Accuracy: $\frac{\text{Correttamente classificati}}{\text{Tutti i campioni}}$

- F1-macro: $\frac{2 * \text{Precision}_M * \text{Recall}_M}{\text{Precision}_M + \text{Recall}_M}$

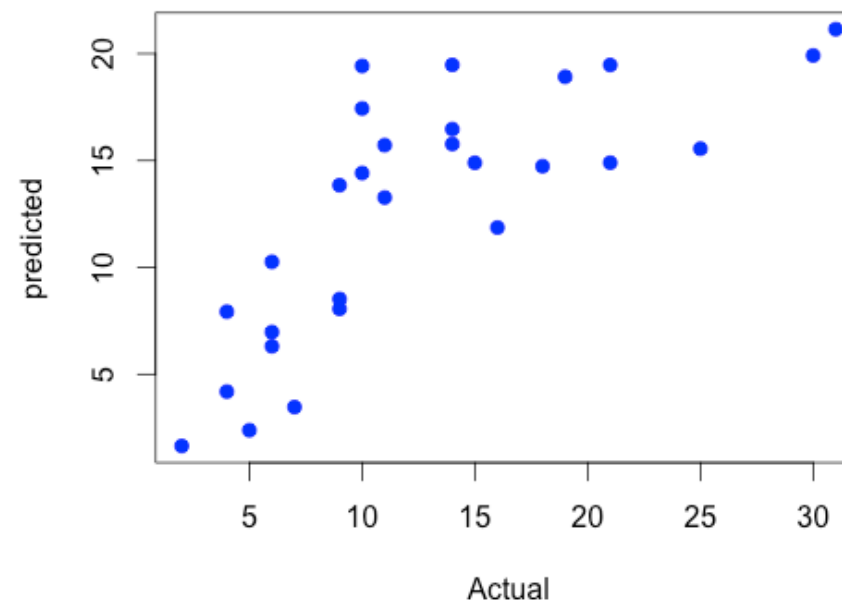
- Confusion Matrix

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

- Variance Score: $\frac{1}{n} \sum (err - \text{mean}(err))^2$

- Mean abs error: $\sum_{i=1}^n \frac{|y_i - x_i|}{n}$

- Predicted vs Real Plot



La metrica per il calcolo del livello di autonomia

Definition 2. A system is autonomous with respect to a set of goals G under a set of outside interventions I (by humans or engineered systems), when the system can achieve all the goals in G , assisted by just the interventions I .

Autonomy and metrics of Autonomy,
Antsaklis P. J.

$$M_i = \frac{n + 1}{5 \cdot n}$$

Livello di autonomia: $L = M_g / M_i$

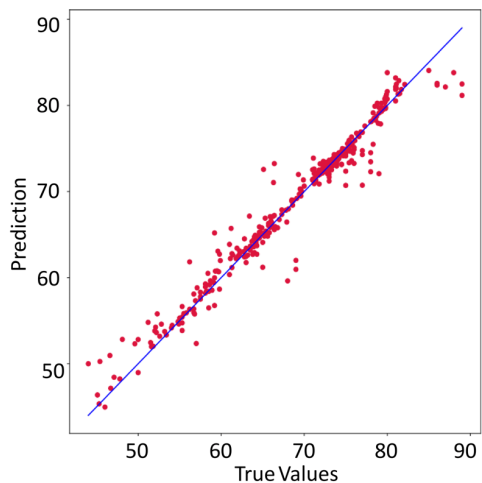
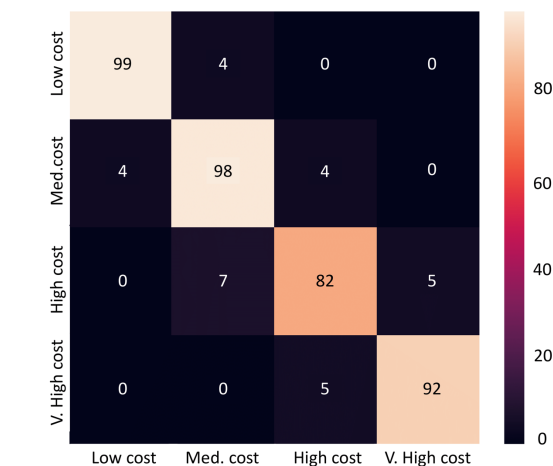
$$M_g = 100 \cdot \sum_{t=1}^n M_t$$

n = numero di esperimenti effettuati, M_t metrica usata per valutare l'esperimento normalizzata a 1

Risultato Finale I

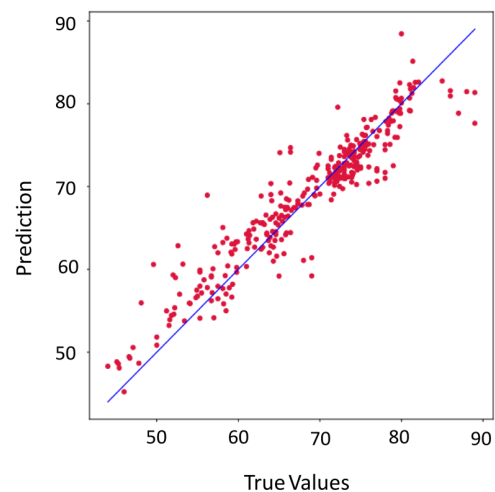
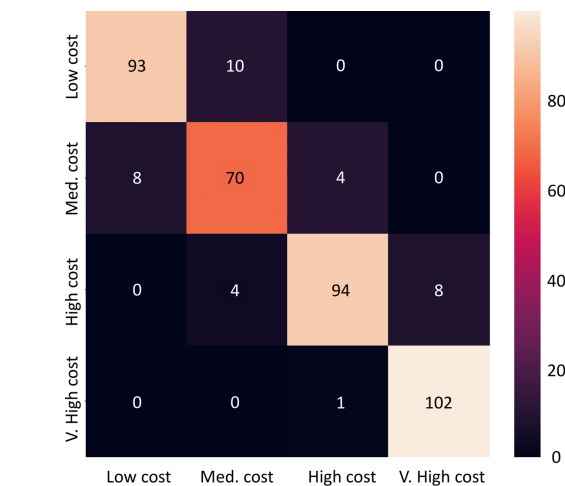
AutoGluon

Liv Autonomia: 87



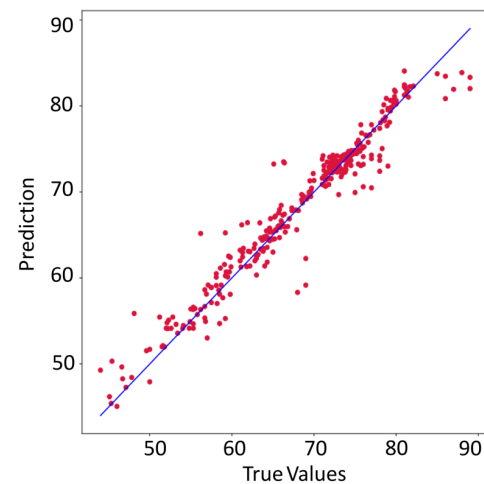
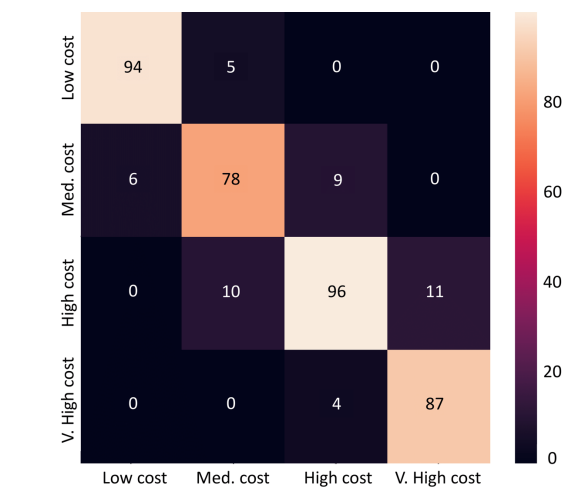
AutoKeras

Liv Autonomia: 74



AutoSkLearn

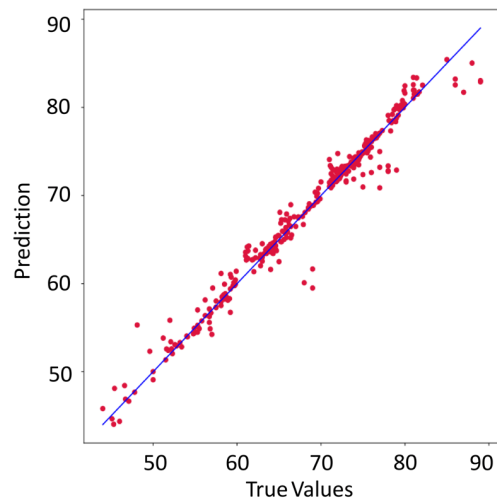
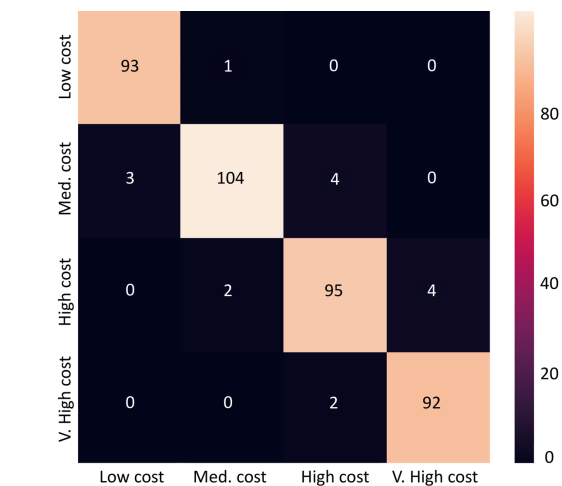
Liv Autonomia: 67



Risultato Finale II

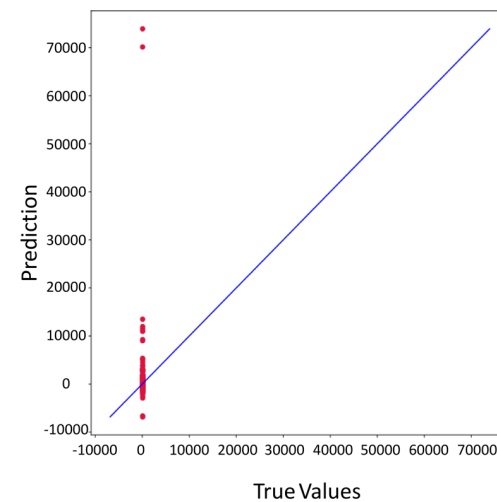
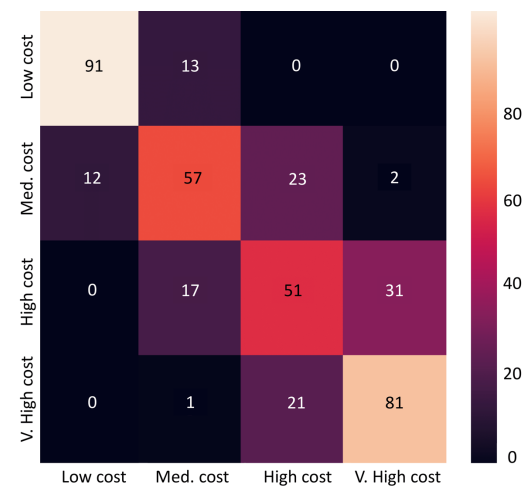
H2O

Liv Autonomia: 59



Ludwig

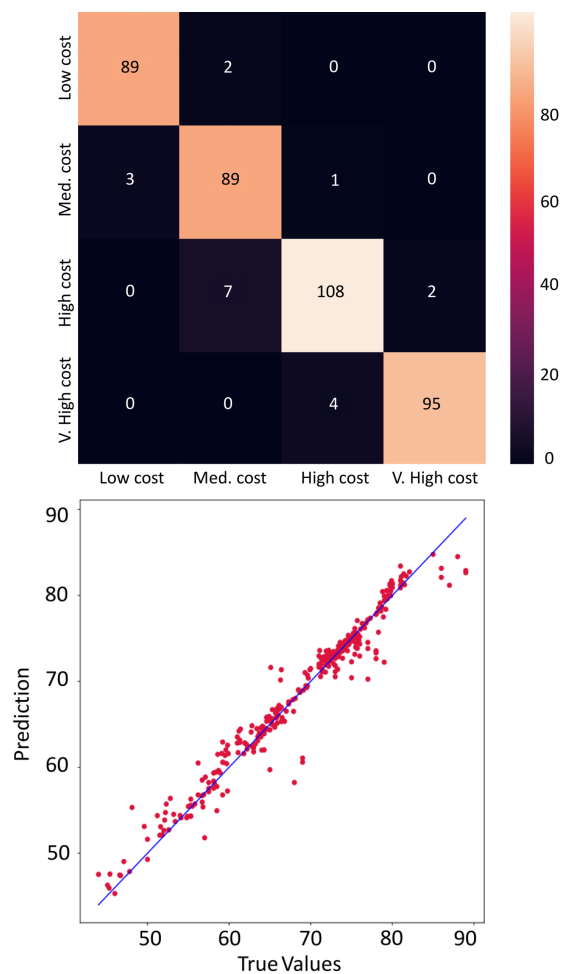
Liv Autonomia: 7



Risultato Finale III

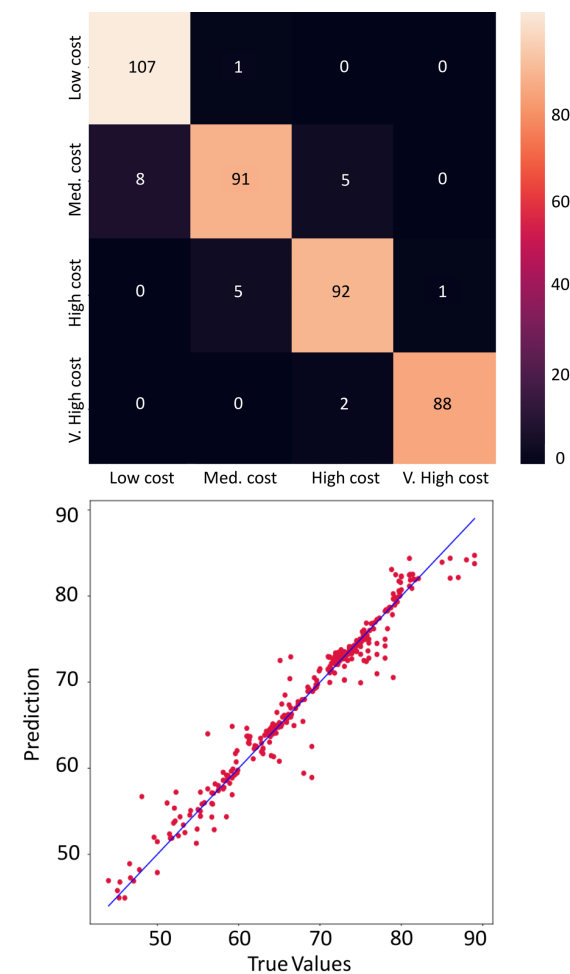
MLJarSupervised

Liv Autonomia: 83



Tpot

Liv Autonomia: 70



AutoML: Machine Learning per tutti?



Task relativamente semplici, ottime prestazioni



Ottimo strumento nelle mani di un esperto



Non ancora pronto per un pubblico non competente

Grazie per l'attenzione

