



DA UN MODELLO PREDITTIVO OPACO AD UN SURROGATO TRASPARENTE: APPLICAZIONE AI PROCESSI DI BUSINESS

Piero Pastore

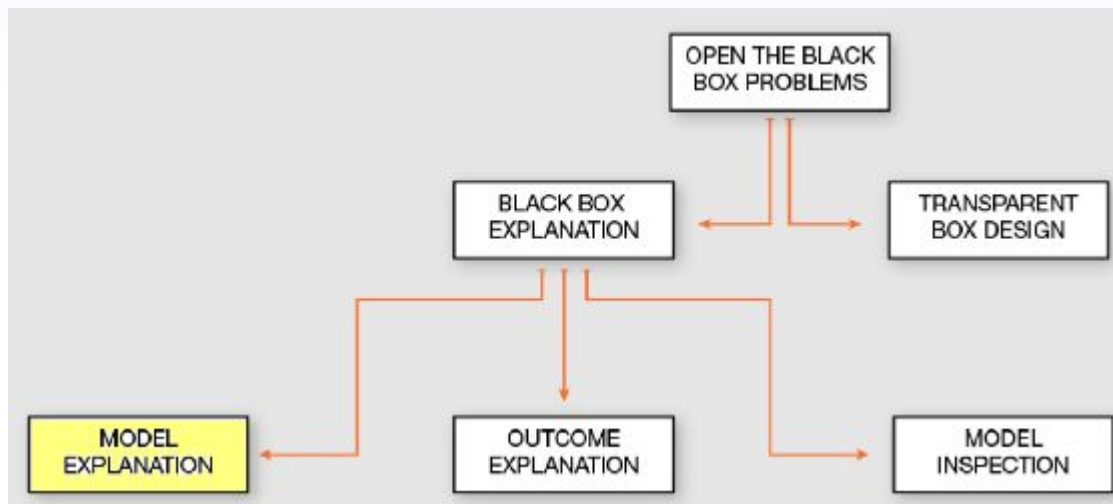
Relatore: Prof. Gabriele Gianini

Anno 2021/2022



CASO DI STUDIO - XAI EXPLAINABLE ARTIFICIAL INTELLIGENCE

XAI e' una nuova branca emergente dell'**IA** che si occupa di **semplificare** e **interpretare** al meglio modelli neurali più complessi.



Tassonomia XAI

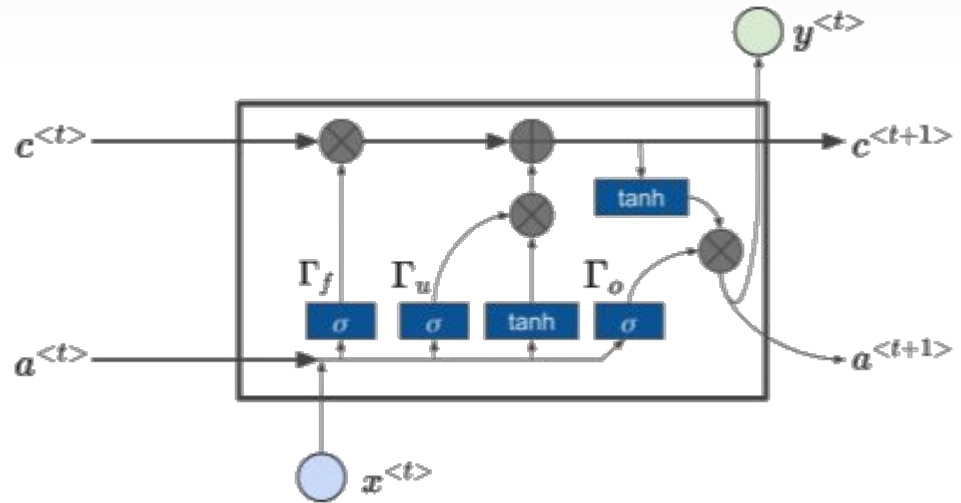
TRASFORMARE UN MODELLO PREDITTIVO OPACO IN UN SURROGATO TRASPARENTE



LONG SHORT TERM MEMORY

LSTM REGRESSOR

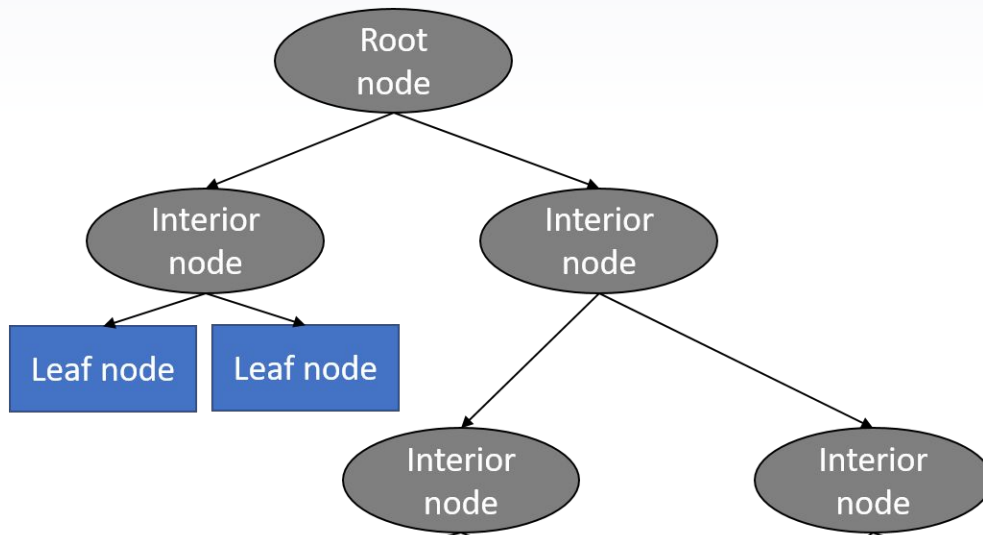
Rete neurale ricorrente che permette di sfruttare le informazioni a lungo termine permettendo di generare previsioni o serie temporali



DECISION TREE

DECISION TREE REGRESSOR

Tecnica basata su una serie di regressori di alberi decisionali che permettono di generare predizioni attraverso un partizionamento ricorsivo binario dei dati



► DATASET

LOG OF LOAN BUSINESS PROCESS DATASET

- ▶ Contiene informazioni su **processi aziendali**, con **task, tipologie di task, costi e utilizzo delle risorse**
- ▶ Fase di **pulizia, formattazione e normalizzazione** prima di essere utilizzato
- ▶ 6879 record

Esempio predizione label

	KIND	TASK	SOURCE	VALUE
t+1	6	1.0	12	55.2
...

TUNING LSTM

Viene utilizzato un criterio di validazione **walk-forward**

- ▶ Ogni **configurazione** del modello viene valutato dopo **10-20 iterazioni**
- ▶ Parametri valutati:

Epoche

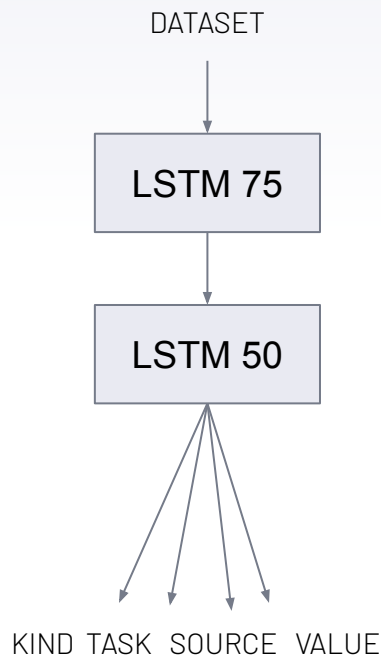
Batch size

- ▶ Iperparametri valutati:

Numero di neuroni

Layer

layer	1,2,3
epoch	[100,150...500]
neurons	[50...200]
batch_size	[4,8...1024]



RISULTATI LSTM

- ▶ **Percentuale accuratezza** dei record
- ▶ Errore stimato mediante **MSE**

	LSTM
TOTALE	92%
SOURCE	92%
TASK	96%
VALUE	88%
KIND	95%

Analisi accuratezza modelli sui dati reali

	LSTM
SOURCE	0.876
TASK	0.002
VALUE	0.921
KIND	0.903

Errore quadratico medio tra i dati reali e i dati stimati dal modello neurale

TUNING DT

Viene utilizzata la tecnica di **grid search** con **cross validation**

- ▶ Modello **allenato** con le predizioni ottenute da LSTM
- ▶ Vengono valutati i seguenti **iperparametri** :

Criterio e scelta del partizionamento

Profondita'

Terminazione albero

criterion	gini, entropy
split	Best, random
max_depth	[12-128]
min_samples_leaf	8-32

RISULTATI DT

- ▶ **Percentuale accuratezza** dei record **validi**
- ▶ Errore stimato mediante **MSE**

	DT
TOTALE	95%
SOURCE	95%
TASK	96%
VALUE	92%
KIND	98%

Analisi accuratezza modello sui dati generati da LSTM

	DT
SOURCE	0.592
TASK	0.002
VALUE	0.602
KIND	0.670

Errore quadratico medio tra i dati iniziali e i dati stimati divisi in base al modello

METRICHE DI VALUTAZIONE

- ▶ **Analisi accuratezza** dei modelli
- ▶ Errore stimato mediante **MSE**
- ▶ Confronto visivo della distribuzione dei valori delle label attraverso **grafici**





CONFRONTI NUMERICI

	LSTM	DT
TOTALE	92%	95%
SOURCE	92%	95%
TASK	96%	96%
VALUE	88%	94%
KIND	95%	96%

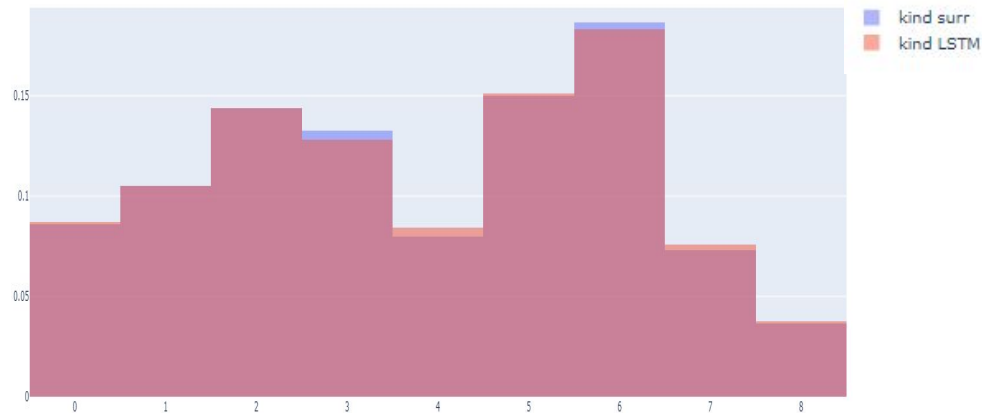
Analisi accuratezza modello neurale e surrogato

	LSTM	DT
SOURCE	0.876	0.592
TASK	0.002	0.002
VALUE	0.921	0.602
KIND	0.903	0.670

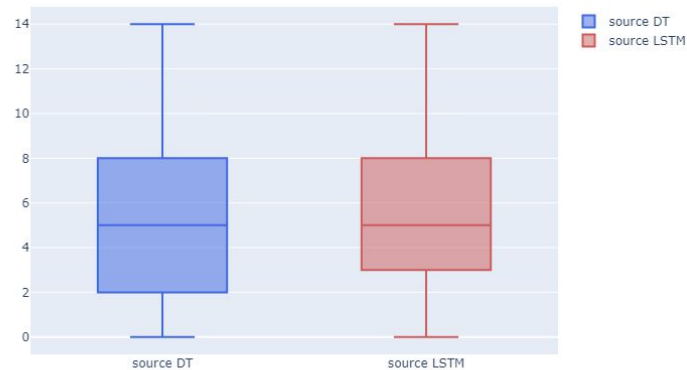
Calcolo errore quadratico medio per entrambe le stime

CONFRONTO GRAFICO

Distribuzione label kind dei valori stimati dai 2 modelli



Variabilita' label source dei valori generati dai 2 modelli



► SVILUPPI FUTURI

- ▶ Utilizzati **nuovi dataset** piu' ricchi ed eterogenei
- ▶ Testare **nuovi modelli trasparenti** che possano semplificare ancora di piu' i risultati ottenuti dall'analisi



**GRAZIE PER
L'ATTENZIONE**