

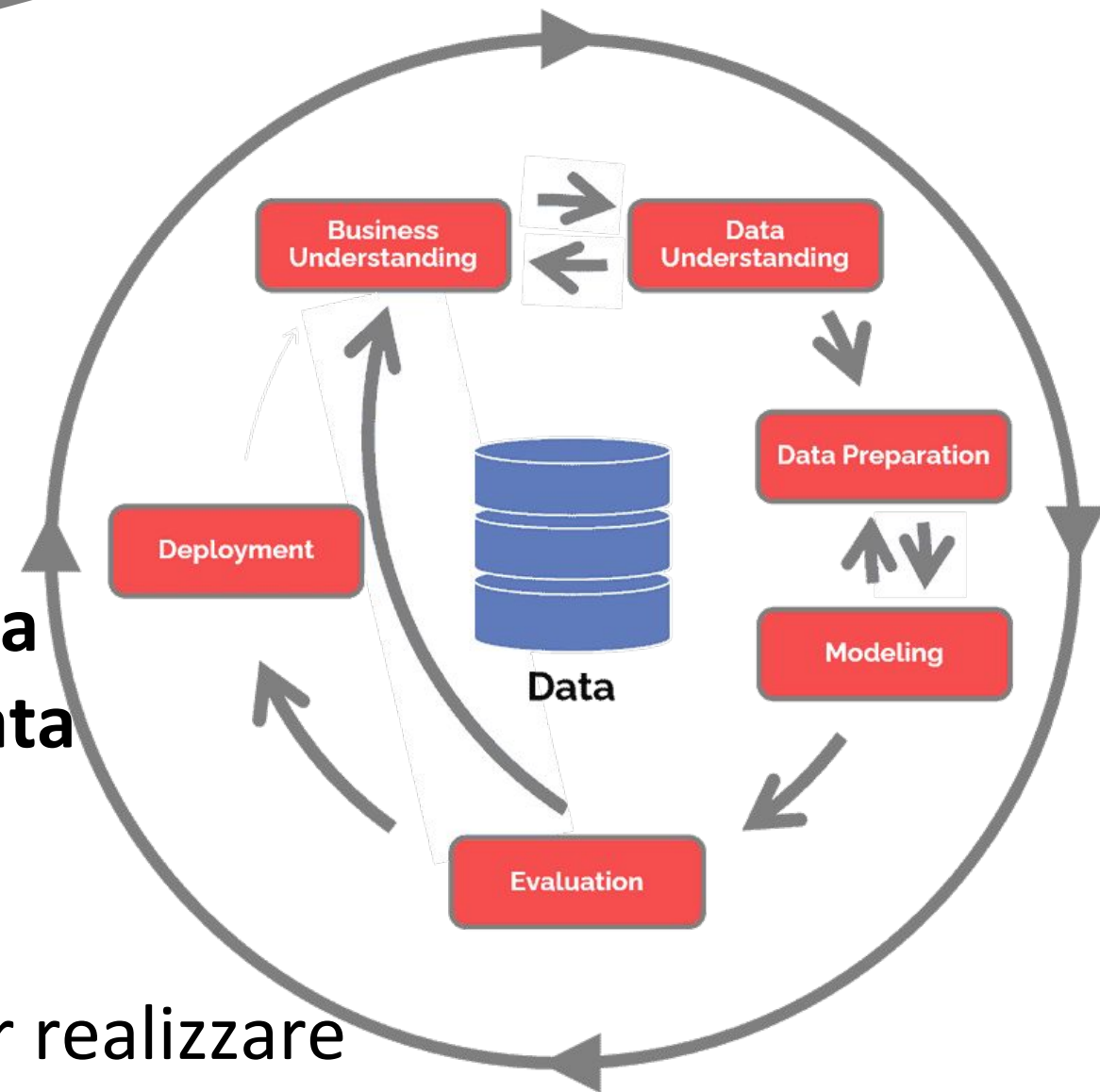


Machine Learning – CRISP-DM

Davide Iacopino

Cosa dobbiamo fare per realizzare un progetto di ML?

- **CRISP-DM** (Cross-Industry Standard Process for Data Mining) è una **metodologia standardizzata e strutturata** per la realizzazione di **progetti di data mining**
- definisce le operazioni da fare per realizzare un progetto di ML, organizzate in 6 fasi
- Processo iterativo e flessibile

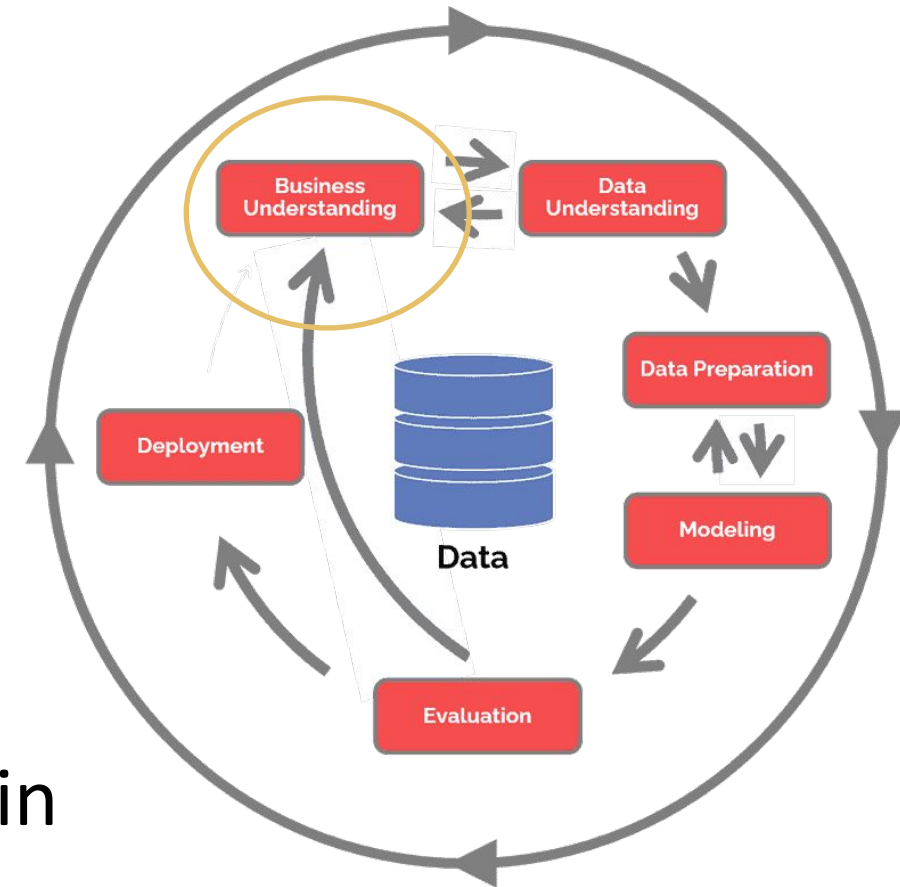


Data mining e Machine Learning

- Il **Data Mining** è il processo di estrazione automatica o semi-automatica di **conoscenza utile, implicita e precedentemente sconosciuta** dai dati.
- Si tratta di informazioni presenti nei dati ma, in qualche modo, nascoste
- Il Machine Learning è la costruzione di modelli in grado (tramite l'apprendimento automatico) di estrarre questa conoscenza dai dati

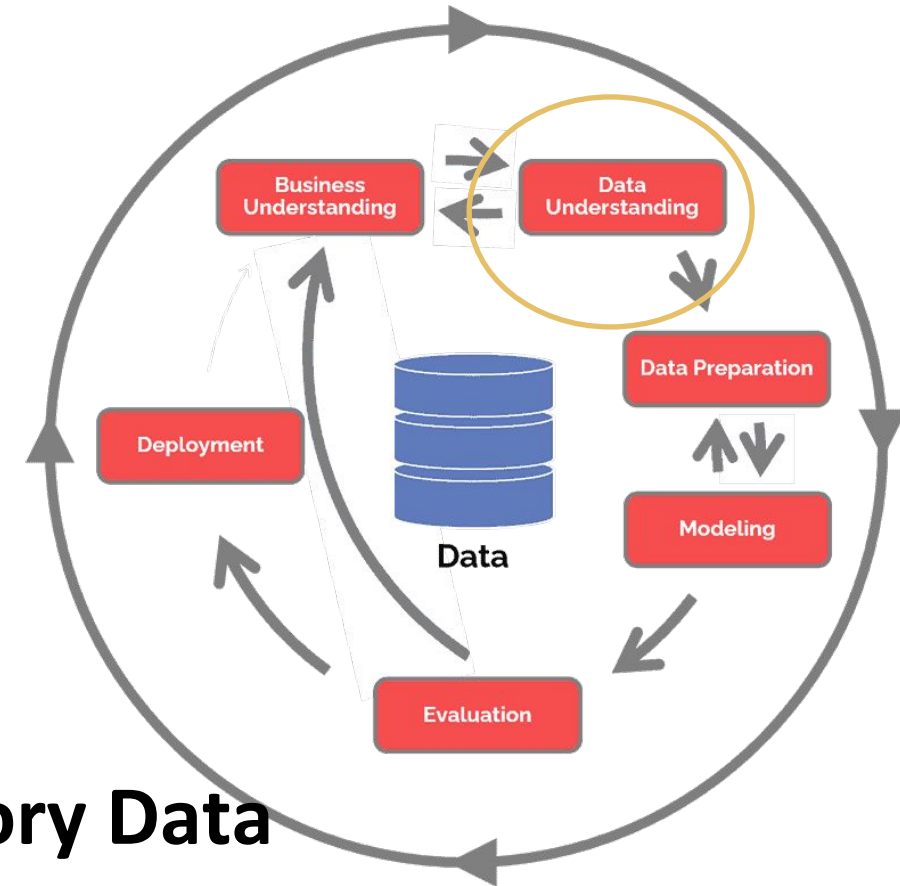
1. Comprensione del dominio

- Definizione degli obiettivi e dei requisiti del progetto, ossia traduzione dei bisogni del cliente in un problema di data mining
- valutazione della situazione attuale (dati a disposizione, dati in fase di raccolta ecc.)
- definizione degli obiettivi di data mining (capire che tipologia di task bisogna affrontare) e pianificazione del progetto



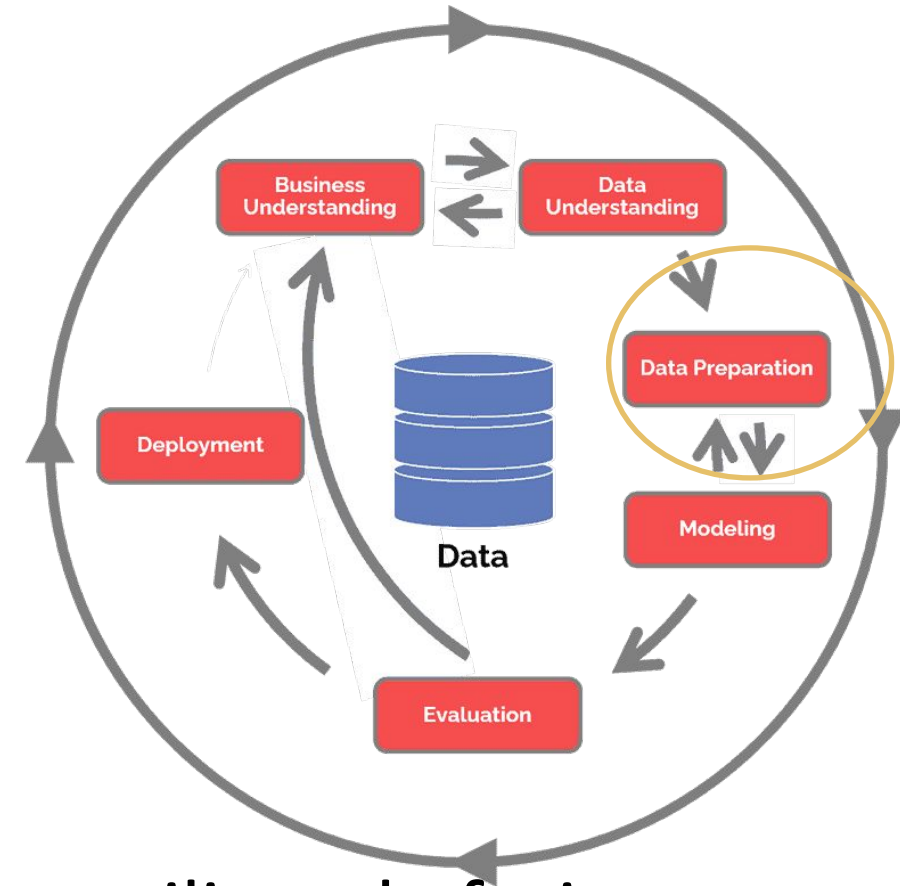
2. Comprensione dei dati

- Esplorazione dei dati (EDA – **Exploratory Data Analysis**) per comprenderli e verificarne caratteristiche e bontà (in termini di quantità, qualità, pulizia ecc).
- Preparazione per la fase successiva, ossia individuazione delle operazioni da fare sui dati



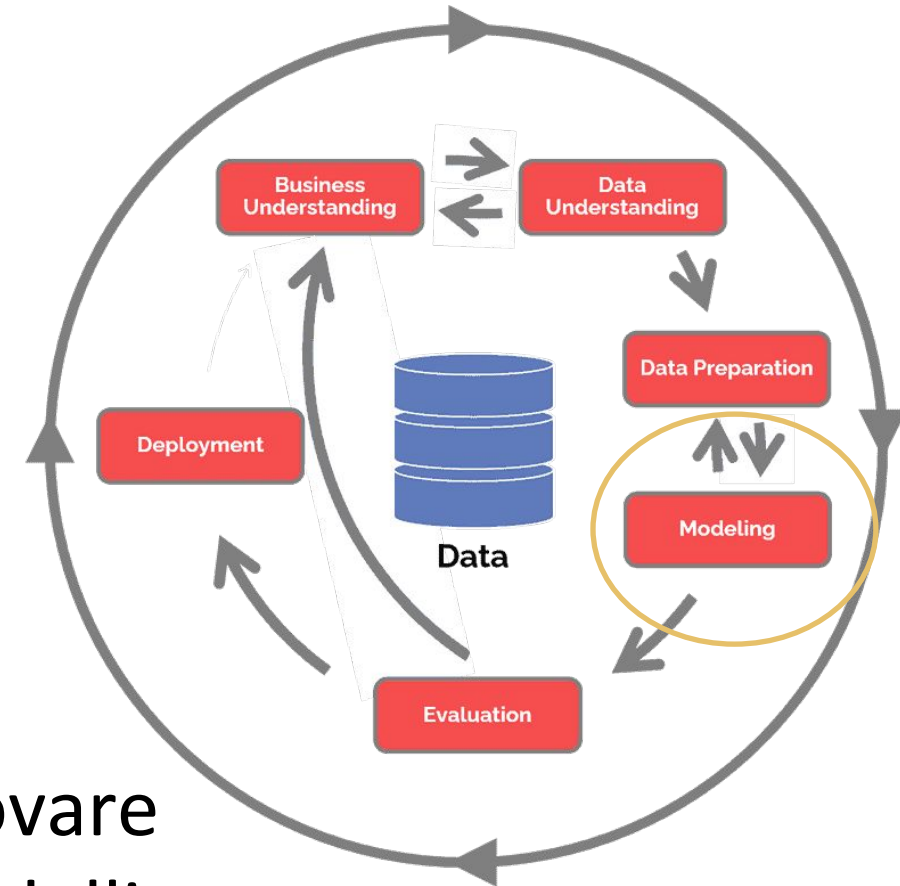
3. Preparazione dei dati

- Integrazione e pulizia dei dati
- Selezione dei dati che ci possono essere utili per le fasi successive e costruzione nuovi dati a partire da quelli disponibili
- Trasformazioni sui dati per renderli più adatti alle operazioni successive



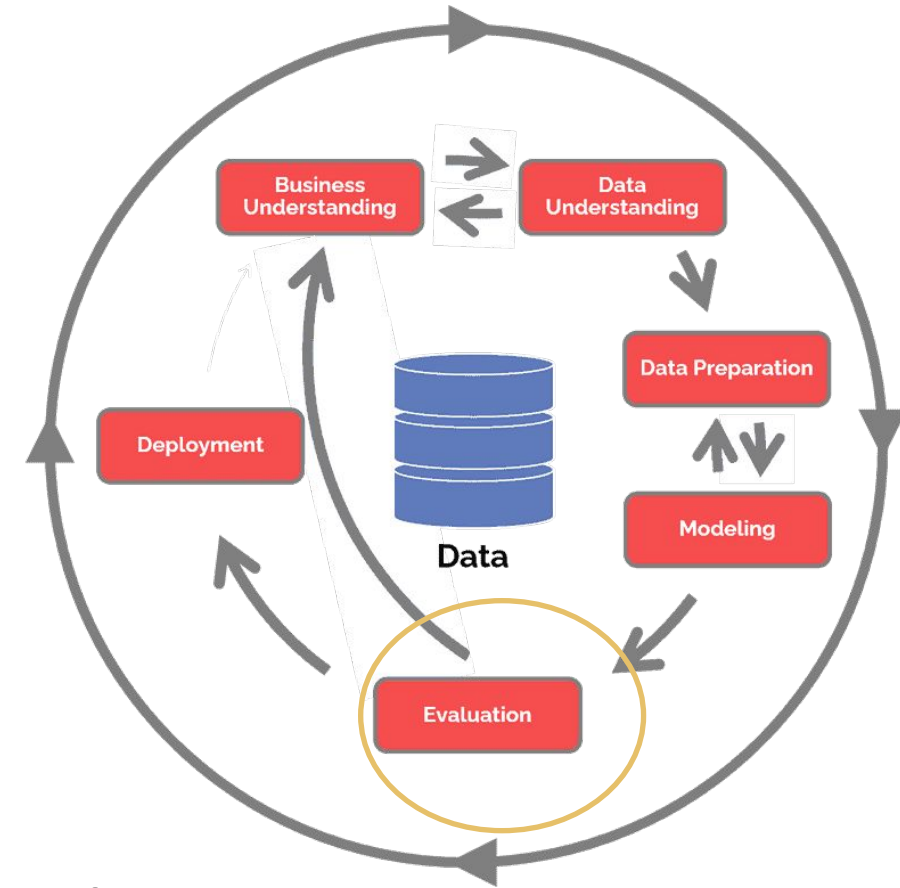
4. Costruzione del modello

- **Selezione** di tecniche e modelli da provare
- **Addestramento** e **valutazione** dei modelli
- Individuazione dei modelli più promettenti
- **Tuning** degli iperparametri dei modelli



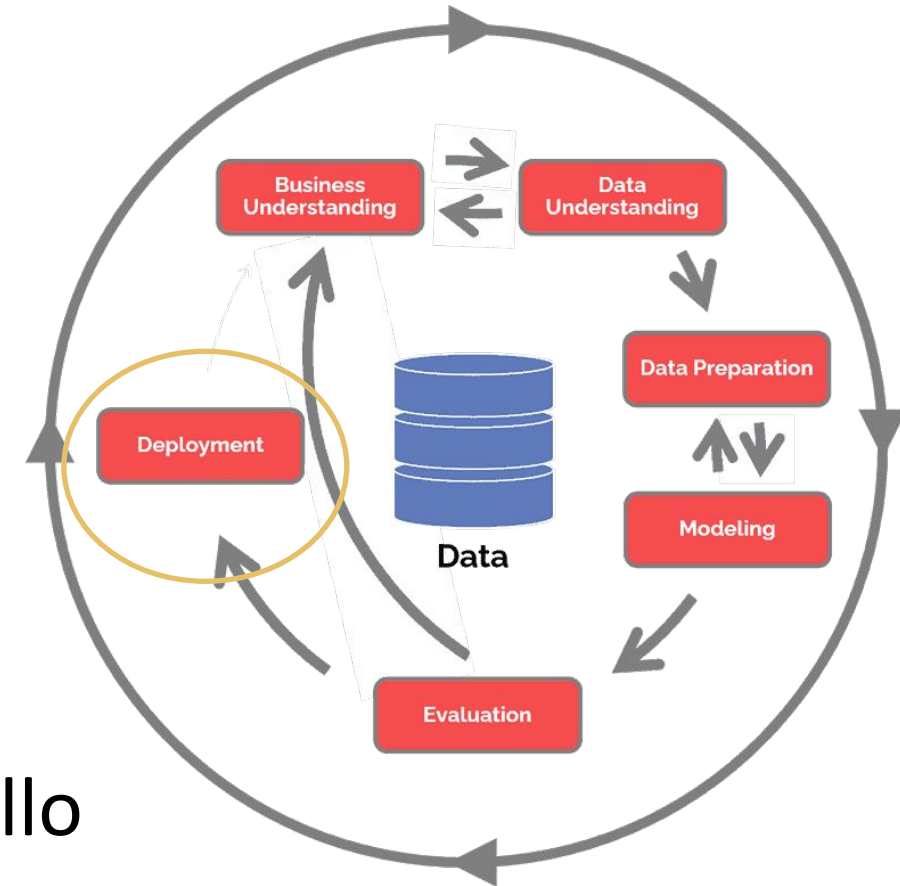
5. Valutazione del modello

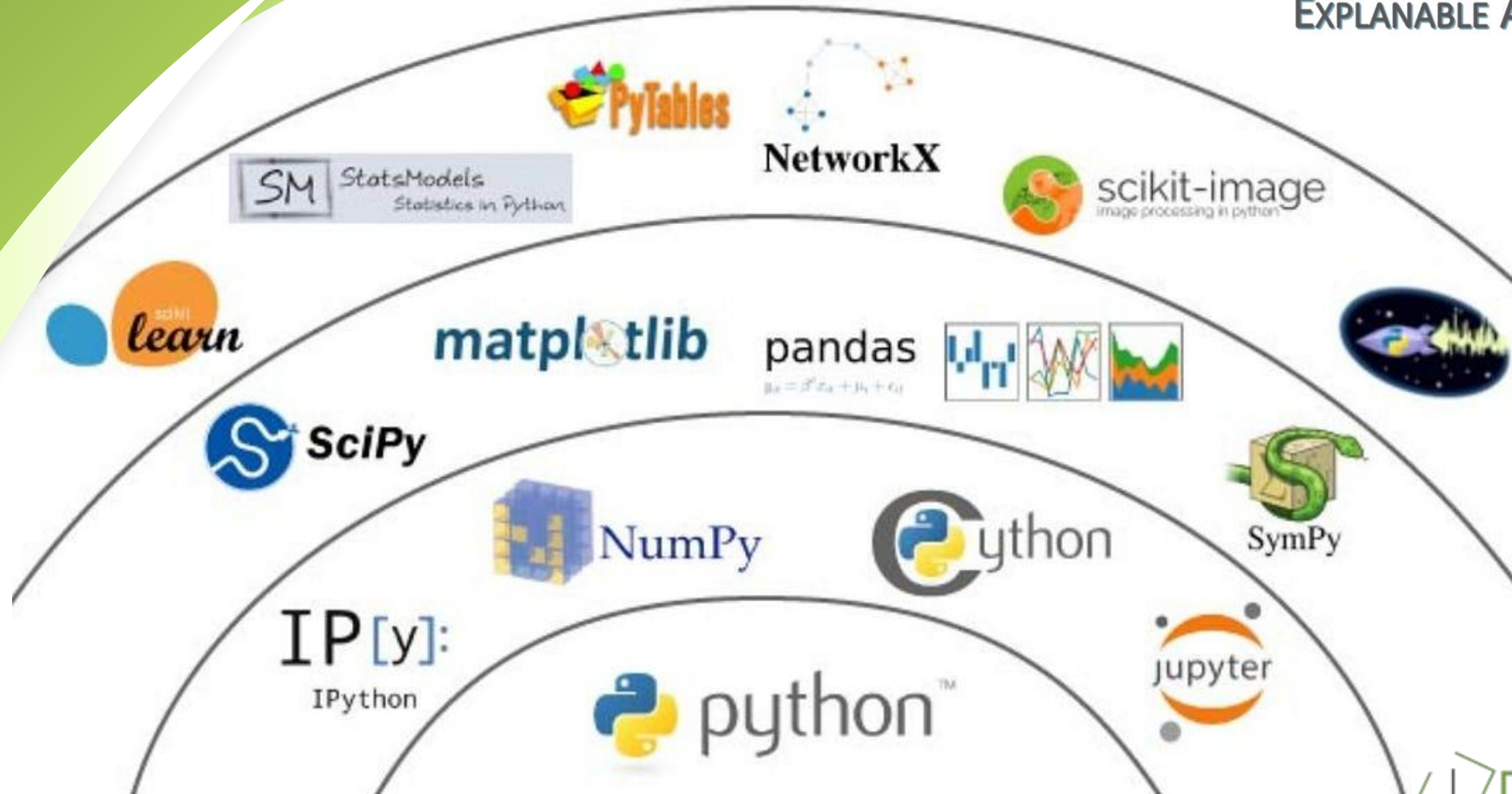
- **Valutazione finale dei risultati** e delle performance del miglior modello ottenuto
- A seconda del modello e del problema che stiamo affrontando, ci sono diverse **misure (metriche) di bontà delle performance**



6. Deployment

- Rilascio e messa in funzione del modello implementando il software per effettuare l'inferenza su nuovi dati
- Spesso i modelli di ML si inseriscono in sistemi software più complessi, quindi in questa fase integriamo il modello col resto del sistema







Salvatore Iiritano

CEO

salvatore.iiritano@revelis.eu

Davide Iacopino

Data Analyst

davide.iacopino@revelis.eu



Rende

V.le della Resistenza, 19/C
87036 Rende (CS)

Parma

Largo L. Mercantini, 13
43125 Parma (PR)



Telefono

(+39) 335.1099492

Fax

(+39) 0984.494269



info@revelis.eu

www.revelis.eu