**Slide**

Una volta trovati tutti i valori per le metriche, non si può essere sicuri al 100% che quei valori siano giusti.

Infatti, le metriche vengono calcolate su elementi del dominio estratti randomicamente e potrebbe darsi che una differente estrazione porti a risultati leggermente diversi.

Quello che si fa in questi casi è calcolare un intervallo di confidenza in cui molto probabilmente è contenuto il reale valore della metrica.

In questo caso noi abbiamo scelto una confidenza del 95%, quindi il reale valore della nostra metrica è contenuto all’interno dell’intervallo che abbiamo trovato, con una probabilità del 95%.

Per quanto riguarda il task di regressione abbiamo calcolato l’intervallo di confidenza dell’errore nella predizione, quindi il valore assoluto della differenza tra la nostra predizione e il groundtruth.

Mentre per quanto riguarda il task di classificazione, abbiamo trovato l’intervallo di confidenza dell’error rate [*e sono rappresentati nella tabella qui sotto*].

**Slide**