

## Descrizione del progetto di ricerca

Il progetto di ricerca, nel quale il lavoro di tesi proposto si inserisce, consta della modellazione, implementazione, addestramento e test di una rete neurale che sia in grado di individuare, elaborando i dati relativi a performance, anzianità di servizio e livello di soddisfazione personale delle risorse professionali, l'insieme di dipendenti maggiormente a rischio di dimissioni nel breve periodo (*employee churn prediction*).

Il componente software risultante dovrà essere reso accessibile come servizio per renderne possibile l'integrazione applicativa con i sistemi gestionali in possesso del dipartimento delle risorse umane.

Il problema della valutazione del rischio dimissioni è classicamente affrontato, in termini di trend percentuali, con metodologie di analisi tipiche del data warehousing. Le reti neurali consentono invece di affrontare la problematica focalizzandosi sui record individuali, stimando il rischio per ciascun dipendente. Tale aspetto è fondamentale considerato che il problema delle dimissioni per le aziende è sostanzialmente legato alla perdita delle competenze e degli investimenti legati alla formazione del personale, per cui il 'costo' percepito non è uniformemente distribuito su tutto il personale (come esempio intuitivo, un tecnico fortemente specializzato, con anzianità di servizio e competenze di nicchia, è più difficile da rimpiazzare di un tecnico junior con competenze generiche).

Una rete neurale addestrata su un numero di occorrenze sufficientemente alto, dovrebbe essere in grado di classificare i dipendenti, a partire dalla loro storia aziendale (in termini di dati quali età, ruolo, stipendio, livello di soddisfazione personale, etc.), in funzione della probabilità di abbandono dell'azienda.

## Obiettivi della tesi

Sarà la realizzazione dei seguenti task di progetto:

1. implementare il modello di rete neurale individuato;
2. pre-elaborare il training set a partire da dati disponibili in letteratura e in funzione delle strutture dati utilizzate dal sistema aziendale di gestione delle risorse umane;
3. addestrare la rete e raffinarne la struttura;
4. pre-elaborare il test set;
5. testare la rete e valutare i risultati ottenuti, in termini di efficacia nella classificazione.

## Tecnologie da utilizzare

- Python 3.7 in ambiente Anaconda.
- Tensor Flow e relative librerie (ad esempio Keras).