

## **Titolo del Progetto:** Named Entity Recognition

### **Obiettivi:**

L'obiettivo principale di questo progetto è sviluppare un sistema di Named Entity Recognition (NER). Gli studenti dovranno:

- Identificare e sfruttare fonti online per acquisire dati rappresentativi di testi contenenti entità nominative da diverse categorie.
- Effettuare un'analisi approfondita dei testi per identificare e categorizzare automaticamente le entità nominative presenti, sviluppando un sistema di NER.
- Implementare un modello in grado di riconoscere automaticamente e accuratamente entità nominative in testi non strutturati.
- Ottimizzare il modello per garantire elevate prestazioni su dati non etichettati.

### **Metodologia di Implementazione:**

- **Selezione delle Fonti Online:** Ricerca e selezione di fonti online, come articoli, blog, o contenuti web, per acquisire dati rappresentativi di testi contenenti entità nominative.
- **Preprocessing e Analisi dei Testi:** Preprocessing dei dati per garantire uniformità nella rappresentazione dei testi.
- **Sviluppo del Modello di NER:** Implementazione di un modello utilizzando algoritmi di machine learning, come Conditional Random Fields o modelli di deep learning come reti neurali ricorrenti o Transformer. Esplorazione di modelli pre-addestrati, come BERT o SpaCy, per migliorare le prestazioni del sistema.
- **Ottimizzazione e Valutazione:** Ottimizzazione del modello per garantire una marcatura accurata delle entità nominative nei testi. Valutazione delle prestazioni del modello attraverso metriche appropriate, come precision, recall, e F1-score.

### **Risultati:**

- Un modello di Named Entity Recognition in grado di identificare e categorizzare automaticamente entità presenti in testi non strutturati.
- Un'analisi dettagliata delle sfide affrontate durante il preprocessing e l'analisi dei testi contenenti entità nominative raccolti online.
- Una valutazione accurata delle prestazioni del modello, compresa la sua capacità di generalizzazione su nuovi dati.
- Un rapporto dettagliato documentante le scelte progettuali, la metodologia implementativa e i risultati ottenuti.