

Junior LLM Engineer – Task Data Extraction

Sono forniti 4 file pdf (KIDs - Key Information Documents) contenenti le caratteristiche anagrafiche di alcuni prodotti finanziari assicurativi.

L'obiettivo del task è implementare un codice Python che attraverso l'utilizzo di LLMs permetta l'estrazione automatica di alcune delle informazioni riportate nei singoli documenti.

Informazioni oggetto dell'estrazione:

1. Campi puntuali:

- ISIN (se e dove presente)
- SRI – indicatore sintetico di rischio (va da 1 a 7)
- RHP – orizzonte di detenzione raccomandato
- NOME DEL PRODOTTO
- NOME DELL'EMITTENTE

2. Campi testuali:

- Target market

3. Campi tabellari (Hint: sarebbe opportuno, prima di interrogare un LLM, estrarre la tabella in maniera strutturata):

- Tutti gli scenari di performance a scadenza in %

Bonus

Creare un'interfaccia utente che permetta l'upload del documento, l'estrazione dei campi e la visualizzazione dei dati estratti.

(per una rapida implementazione si consiglia l'utilizzo di <https://streamlit.io/>)

Durante l'intervista verrà discusso l'approccio utilizzato e sarà testato il codice prodotto.

Non ci sono vincoli sulle librerie che il candidato potrà utilizzare.

Esempio di output atteso (documento KID1.pdf):

- ISIN: LU0690021539
- SRI: 2
- RHP: 5
- NOME DEL PRODOTTO: Unit Personal Private
- NOME DELL'EMITTENTE: Fineco Asset Management
- TARGET MARKET: il Fondo Esterno è rivolto ai clienti aventi una buona conoscenza e/o esperienza dei mercati finanziari e assicurativi e una bassa tolleranza al rischio finanziario, i quali intendono investire il proprio capitale nel medio periodo accettando i possibili rischi di perdita connessi all'investimento finanziario.
- SCENARI DI PERFORMANCE A RHP:
 - stress -5.82%,
 - sfavorevole -2.73%,
 - moderato 0.13%.
 - favorevole 0.99%