

MATCHER

Matcher è uno strumento in grado di **automatizzare il processo di pulizia dei tuoi dati**, andando ad individuare le righe duplicate tramite algoritmi di machine learning. Si tratta di una soluzione rapida ed efficace per andare a migliorare la qualità del dato, identificando ad esempio all'interno delle anagrafiche i record appartenenti alla stessa entità.

A differenza dei sistemi tradizionali di deduplica, Matcher si serve di un algoritmo di machine learning in grado di identificare non solo i record identici, ma anche record molto simili associabili alla stessa entità.

Nome	Cognome	Città	Età
Alberto	Rossi	Milano	30
Luigi	Bianchi	Parma	50
Rossi	Alberto	Milano	30
Anna	Belli	Roma	32
Rossella	Nuovo	Torino	44
Anna	Bella	Roma	32
Luigi	Bianchi	Parma	51
Paola	Giorni	Milano	36



Nome	Cognome	Città	Età
Alberto	Rossi	Milano	30
Luigi	Bianchi	Parma	51
Anna	Belli	Roma	32
Rossella	Nuovo	Torino	44
Paola	Giorni	Milano	36

UTILIZZO

Questo strumento è utile per due tipologie principali di operazioni:

- Eliminazione di duplicati
- Cross-matching tra tabelle che non hanno nessuna chiave identificativa in comune

BENEFICI E PECULIARITA'

- Il tool è applicabile a qualsiasi tipologia di file o base dati
- I tempi di adozione sono ridotti, grazie ad un training del modello rapido e preciso
- L'architettura può essere scalata per lavorare con grandi moli di dati

INPUT

Per testare l'efficienza del prodotto occorre solamente fornire un file di qualche centinaio di righe, tra cui alcune ripetizioni.

OUTPUT

Il modello identifica gruppi di istanze praticamente identiche, all'interno della base dati. Indicando quale tra le istanze è quella corretta, la base dati viene automaticamente deduplicata mantenendo la versione corretta di ciascuna istanza.