

REFERENCA: DEDUPLIKIMI I TE DHENAVE

CFARE ESHTE DEDUPLIKIMI?

Deduplikimi eshte procesi i identifikimit dhe eliminimit te rekordeve te dyfishta ne nje databaze. Rekorde duplike krijohen kur i njejtë entitet (person, biznes, adresë) futet me te dhena paksa te ndryshme.

PSE ESHTE I RENDESHISHEM?

- * Statistika te sakta - numri real i qytetareve/sherbimeve
- * Kursim kostosh - shmangje e sherbimeve te dyfishta
- * Cilesie e te dhenave - databaze e paster dhe e besueshme
- * Perputhje ligjore - GDPR kerkon saktesi te dhenash

TEKNIKAT E DEDUPLIKIMIT

Teknika	Pershkrimi	Perdorimi
Exact Match	Perputhje e plete e fushave	Duplike identike
Fuzzy Match	Perputhje e perafert	Gabime shkrimi
Phonetic Match	Tingellim i njashem	Variante emrash
Token Match	Fjale te njejtë ne rend tjeter	Emra te invertuar

ALGORITMET E NGJASHMERISE

Algoritmi	Si Funksionon
Levenshtein	Numeron ndryshimet minimale (insert/delete/replace)
Jaro-Winkler	Favorizon perputhjet ne fillim te stringes
Soundex	Kodon tingellimin fonetik te fjalet
N-gram	Krahason sekuencia karakteresh

PRAKTIKAT ME TE MIRA

- Standardizoni te dhenat para kahasimit (kaptalizim, formate)
- Perdorni shume fusha per kahasim, jo vetem emrin
- Vendosni pragje te ndryshme per fusha te ndryshme
- Rishikoni manualisht perpara bashkimit
- Mbani log te te gjitha bashkimeve per auditim

6. Beni backup para cdo operacioni bashkimi

Referenca Deduplikimi - Workshop AI Biometrike