



encrypt

A scalable and practical
privacy-preserving framework

Pre-Processing Tool

Advanced Data Pre-Processing and PII Extraction Tool

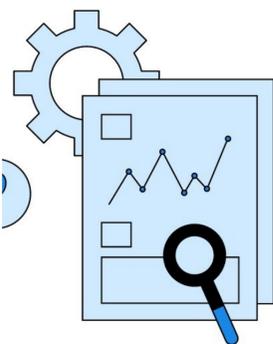


- Tool pre-processes raw datasets and extracts PII across key domains
- Supports Privacy-Preserving Tools such as FHE, DP, TEE
- Cleans, formats and transforms data
 - Minimize computational overhead and improves privacy
- Uses Regular Expressions, NEPRII Python Library, BERT-based NER

OBJECTIVES

Goal: Prepare datasets for secure, standards-compliant analytics

Step	Technique	Purpose
Remove NaNs	Clean dataset	Ensures data usability
Clean symbols (e.g., "-", "14/3")	Normalize data types	Reduces dimension explosion in FHE
Handle missing values	Data imputation or removal	Improves accuracy
Remove duplicates	Optimize size	Enhances performance
Convert categorical to numeric	One-hot encoding or label encoding	FHE compatibility
Feature selection	Remove correlated features	Minimize dimensions for FHE



RAW DATA



Outputs used for

- Knowledge Graphs
- Recommendation Engine
- FHE, DP, TEE modules



Funded by
the European Union

This work is supported by the European Union's Horizon Europe programme under grant agreement No 101070670.

Disclaimer: Views and opinions expressed are those of the ENCRYPT consortium authors only and do not necessarily reflect those of the European Union or its delegated Agency DG CNECT. Neither the European Union nor the granting authority can be held responsible for them.