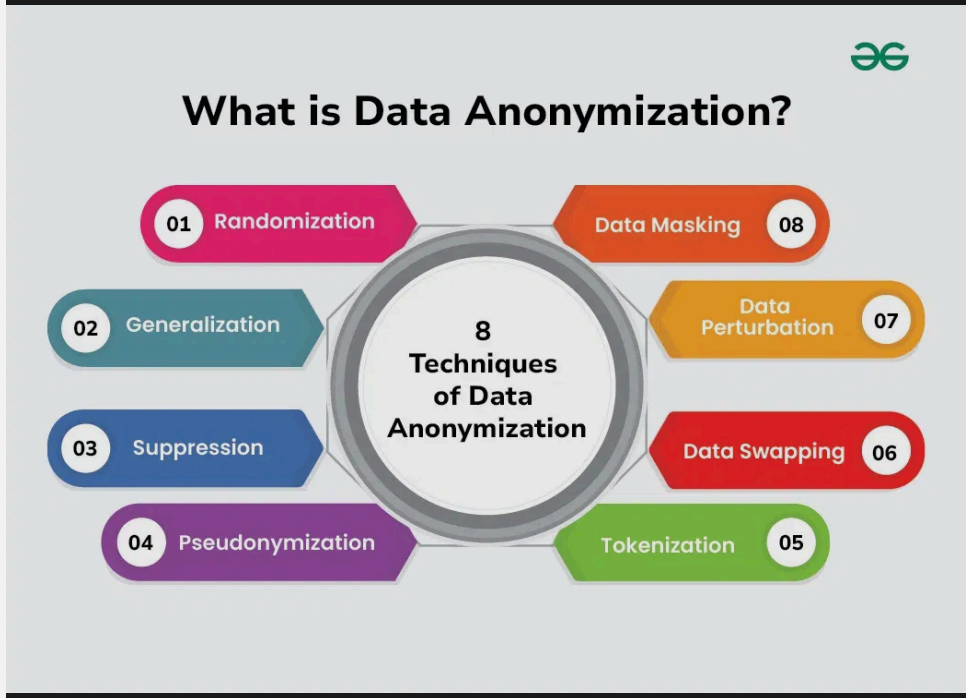
POLISH DATA ANONYMIZATION

Fine-tuned HerBERT NER model
for 25 sensitive classes.

- 25 Klas (RODO)
- 100% Offline
- Brak zewnętrznych API
- Wykorzystanie HerBER



DANE I PRZYGOTOWANIE

ŹRÓDŁO I FORMAT

Wejście: `orig.txt` (raw) + `anonymized.txt` (gold).

Konwersja: `convert_original_data.py` → CoNLL.

Zakres: 25 klas wrażliwych (PESEL, adres, imię, itp.).

PRZETWARZANIE

Tokenizer: HuggingFace (HerBERT).

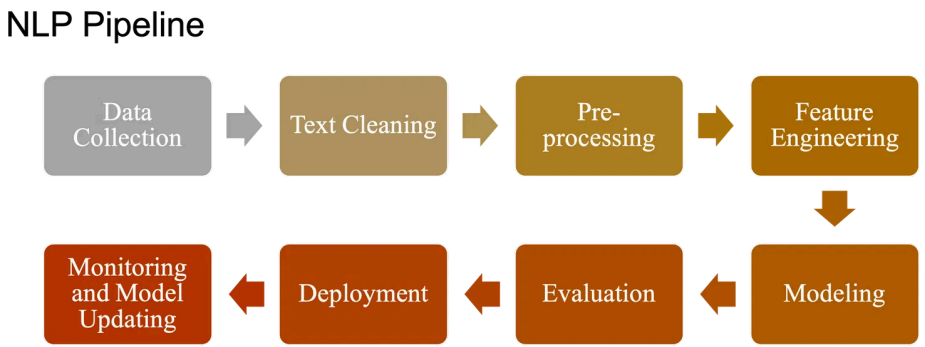
Split: Train / Dev / Test.

Max Length: 256 (trening), 512 (inference).

CEL

Przygotowanie danych treningowych dla modeli PL (PLLuM).

Zachowanie struktury gramatycznej i sensu.



MODEL I TRENING

MODEL BAZOWY

allegro/herbert-base-cased

Polski BERT zoptymalizowany językowo.

FINE-TUNING NER

BIO Tagging Scheme

51 etykiet (B-/I- dla 25 klas + O).

PROCES TRENINGOWY

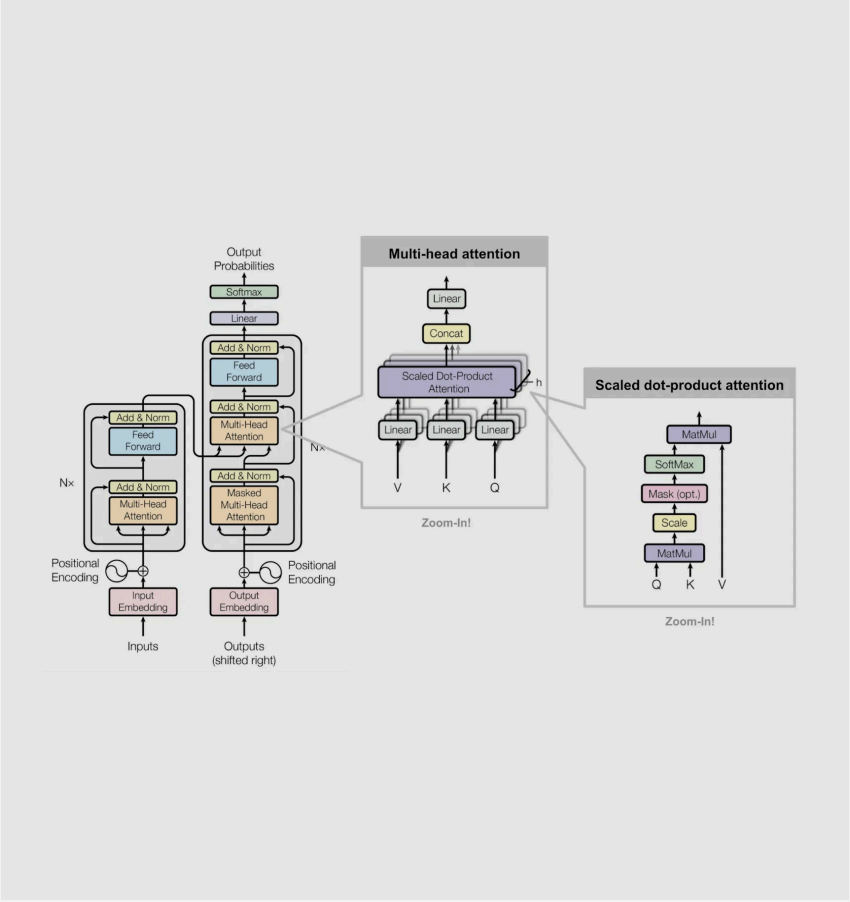
Seqeval F1 Metric

Skrypt: `train_herbert_ner.py train`

CHECKPOINTING

Best Model Saving

Lokalny zapis: `model_output_finetuned/`

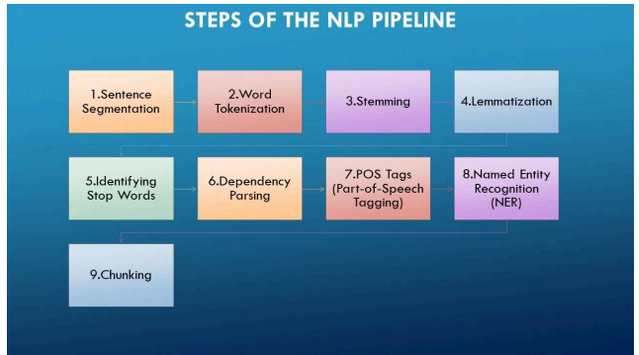


INFERENCJA I EWALUACJA

PIPELINE INFERENCYJNY

```
train_herbert_ner.py predict --input orig.txt
```

1. **Tokenizacja:** Podział na subtokeny HerBERT.
2. **Predykcja:** Model przewiduje tagi BIO.
3. **Offset Mapping:** Mapowanie na znaki oryginału.
4. **Scalanie:** Łączenie encji (B- + I-).
5. **Anonimizacja:** Wstawienie tagów [PESEL].



METRYKI JAKOŚCI

```
evaluate_anonymization.py --gold anonymized.txt
```

PRECISION
Dokładność detekcji (TP/TP+FP)

RECALL
Pokrycie encji (TP/TP+FN)

F1 SCORE
Średnia harmoniczna (Micro/Macro)

PERFECT LINES
Liczba w pełni poprawnych zdań

Output: Raport CSV + Log błędów.
Hardware: Auto-detect CUDA / MPS / CPU.