

NAMED ENTITY RECOGNITION

- Abhi Bedi
- Rishabh Shukla

Problem We are Solving?

We are using eBay listing titles for NER. A few examples of NER labeling of listing titles are shown below (these examples are in English to illustrate the concept, the challenge data will have German language listing titles).



About Data

- 10 million randomly selected unlabeled item titles from eBay Germany: "Athletic Shoes" categories.
 - 10,000 labeled item titles provided.
- Context
 - "New" tagged as "No Tag" in "New shoes", but "Marke" (brand) in "New Balance".
- Misspellings
- No Tag
 - Punctuation, and words adding no meaning (black & white), prepositions.
- Obscure
 - Undeciphered

What is NER?

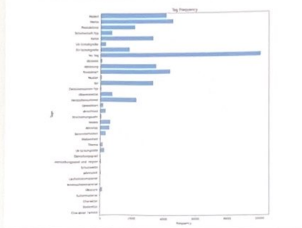
Named entity recognition (NER) is a fundamental task in Natural Language Processing (NLP) and one of the first stages in many language understanding tasks.

Pre-Processing Steps

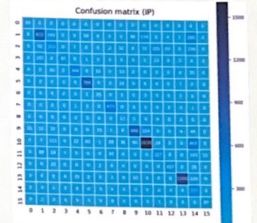
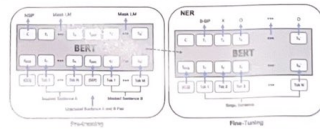
- Tokenization
- Attention Mask (Specific to Bert)
- Padding/Truncation
- Label Encoding
- BERT Specific Input Format

[illegible]

Data Visualization



Modeling with BER



Cross entropy loss is given by

$$\ell(x, y) = L = \{l_1, \dots, l_N\}^T, \quad l_n = -w_n \log \frac{\exp(x_{n,y})}{\sum_{c=1}^C \exp(x_{n,c})}.$$

Accuracy 55%-65%

