

ML Engineer Assignment for Soleadify

by Olha Babich, your future ML Engineer :)

Task

We have to create an ML model that is able to extract products from Furniture Stores.

Subtasks

1

**Somehow
get data for
model
training**

2

**Make some
data
preparation**

3

**Train a
model**

4

**Evaluate
the model**

Data retrieval

- Check if the site works
- Check if it has a sitemap
- Gather all products with links from the sitemap
- Scrape product description

Data preparation

- Concatenate product name with description
- Annotate data using IOB-tagging scheme, considering that description does not have any named entity

Model training

- Create Datasets and DataLoaders
- Define the model
(BertForTokenClassification)
and load it with the
pretrained weights
- Train the model

Model results

- For model evaluation
seqeval library is used

precision	recall	f1-score	n_elements
0.77	0.81	0.79	225

More detail on full doc