

Her göreve özel kullanılacak Simple Transformers modeli, kullanıcının modeli kendi kullanım durumlarına göre kolayca uyarlamasını sağlamak için tonlarca yapılandırma seçeneğiyle birlikte gelir. "<https://simpletransformers.ai/docs/usage/> "

```
!pip install simpletransformers
```



Code snippets

R



Want more memory
and disk space?



[Upgrade to Colab Pro](#)

Not connected to runtime.

[Manage sessions](#)

[Change runtime type](#)

Looking in indexes: <https://pypi.org/simple>, [https://](https://pypi.org/simple)
Collecting simpletransformers
 Downloading simpletransformers-0.63.7-py3-none-any
 |██| 249 kB 16.4 MB/s
Requirement already satisfied: regex in /usr/local/lib/python3.8/site-packages (2020.10.11)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.8/site-packages (0.22.2)
Collecting datasets
 Downloading datasets-2.4.0-py3-none-any.whl (365 kB)
 |██| 365 kB 69.2 MB/s
Collecting segeval
 Downloading segeval-1.2.2.tar.gz (43 kB)
 |██| 43 kB 2.6 MB/s
Requirement already satisfied: pandas in /usr/local/lib/python3.8/site-packages (1.1.5)
Requirement already satisfied: tqdm>=4.47.0 in /usr/local/lib/python3.8/site-packages (4.62.3)
Requirement already satisfied: tensorboard in /usr/local/lib/python3.8/site-packages (2.4.1)
Requirement already satisfied: tokenizers in /usr/local/lib/python3.8/site-packages (0.10.1)
Collecting wandb>=0.10.32
 Downloading wandb-0.13.1-py2.py3-none-any.whl (1.8 MB)
 |██| 1.8 MB 60.8 MB/s
Collecting sentencepiece
 Downloading sentencepiece-0.1.97-cp37-cp37m-manylinux1_x86_64.whl (1.3 MB)
 |██| 1.3 MB 54.0 MB/s
Requirement already satisfied: requests in /usr/local/lib/python3.8/site-packages (2.25.1)
Requirement already satisfied: scipy in /usr/local/lib/python3.8/site-packages (1.5.4)
Requirement already satisfied: transformers>=4.6.0 in /usr/local/lib/python3.8/site-packages (4.10.0)
Requirement already satisfied: numpy in /usr/local/lib/python3.8/site-packages (1.20.1)
Collecting streamlit
 Downloading streamlit-1.12.0-py2.py3-none-any.whl (9.1 MB)
 |██| 9.1 MB 64.0 MB/s
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.8/site-packages (21.3)
Requirement already satisfied: filelock in /usr/local/lib/python3.8/site-packages (3.0.12)
Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.8/site-packages (4.8.1)
Requirement already satisfied: huggingface-hub<1.0, >=0.5.0 in /usr/local/lib/python3.8/site-packages (0.8.1)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.8/site-packages (5.4.1)
Requirement already satisfied: typing-extensions>=3.7.4 in /usr/local/lib/python3.8/site-packages (4.1.1)
Requirement already satisfied: pyparsing!=3.0.5, >=2.4.2 in /usr/local/lib/python3.8/site-packages (3.0.7)
Collecting sentry-sdk>=1.0.0
 Downloading sentry_sdk-1.9.5-py2.py3-none-any.whl (157 kB)
 |██| 157 kB 67.7 MB/s
Requirement already satisfied: promise<3, >=2.0 in /usr/local/lib/python3.8/site-packages (2.3)
Requirement already satisfied: protobuf<4.0dev, >=3.12.0 in /usr/local/lib/python3.8/site-packages (3.17.3)
Requirement already satisfied: six>=1.13.0 in /usr/local/lib/python3.8/site-packages (1.16.0)
Requirement already satisfied: setuptools in /usr/local/lib/python3.8/site-packages (57.5.0)
Collecting shortuuid>=0.5.0
 Downloading shortuuid-1.0.9-py3-none-any.whl (9.4 kB)
Collecting GitPython>=1.0.0
 Downloading GitPython-3.1.27-py3-none-any.whl (181 kB)
 |██| 181 kB 73.5 MB/s
Collecting setproctitle
 Downloading setproctitle-1.3.2-cp37-cp37m-manylinux1_x86_64.whl (28 kB)
Requirement already satisfied: Click!=8.0.0, >=7.0 in /usr/local/lib/python3.8/site-packages (7.1.2)
Requirement already satisfied: psutil>=5.0.0 in /usr/local/lib/python3.8/site-packages (5.9.0)
Collecting docker-pycreds>=0.4.0
 Downloading docker_pycreds-0.4.0-py2.py3-none-any.whl (10 kB)
Collecting pathtools
 Downloading pathtools-0.1.2.tar.gz (11 kB)
Collecting gitdb<5, >=4.0.1
 Downloading gitdb-4.0.9-py3-none-any.whl (63 kB)
 |██| 63 kB 2.1 MB/s
Collecting smmap<6, >=3.0.1

```
Downloading smmap-5.0.0-py3-none-any.whl (24 kB)
Requirement already satisfied: certifi>=2017.4.17 in
Requirement already satisfied: chardet<4,>=3.0.2 in
Requirement already satisfied: urllib3!=1.25.0,!>=1.2
Requirement already satisfied: idna<3,>=2.5 in /usr/
Collecting sentry-sdk>=1.0.0
  Downloading sentry_sdk-1.9.4-py2.py3-none-any.whl
    |████████████████████████████████████████| 157 kB 77.9
  Downloading sentry_sdk-1.9.3-py2.py3-none-any.whl
    |████████████████████████████████████████| 157 kB 74.6
  Downloading sentry_sdk-1.9.2-py2.py3-none-any.whl
    |████████████████████████████████████████| 157 kB 68.2
  Downloading sentry_sdk-1.9.1-py2.py3-none-any.whl
    |████████████████████████████████████████| 157 kB 61.8
  Downloading sentry_sdk-1.9.0-py2.py3-none-any.whl
    |████████████████████████████████████████| 156 kB 57.5
Requirement already satisfied: dill<0.3.6 in /usr/loc
Collecting multiprocessing
  Downloading multiprocessing-0.70.13-py37-none-any.whl
    |████████████████████████████████████████| 115 kB 70.0
Collecting xxhash
  Downloading xxhash-3.0.0-cp37-cp37m-manylinux_2_17
    |████████████████████████████████████████| 212 kB 71.4
Requirement already satisfied: aiohttp in /usr/local
Requirement already satisfied: fsspec[http]>=2021.11
Collecting responses<0.19
  Downloading responses-0.18.0-py3-none-any.whl (38
Requirement already satisfied: pyarrow>=6.0.0 in /us
Collecting urllib3!=1.25.0,!>=1.25.1,<1.26,>=1.21.1
  Downloading urllib3-1.25.11-py2.py3-none-any.whl (
    |████████████████████████████████████████| 127 kB 65.5
Requirement already satisfied: charset-normalizer<3.0
Requirement already satisfied: async-timeout<5.0,>=4
Requirement already satisfied: aiosignal>=1.1.2 in /
Requirement already satisfied: yarl<2.0,>=1.0 in /us
Requirement already satisfied: attrs>=17.3.0 in /usr
Requirement already satisfied: multidict<7.0,>=4.5 in
Requirement already satisfied: frozenlist>=1.1.1 in
Requirement already satisfied: asyncctest==0.13.0 in
Requirement already satisfied: zipp>=0.5 in /usr/loc
Requirement already satisfied: python-dateutil>=2.7.
Requirement already satisfied: pytz>=2017.3 in /usr/
Requirement already satisfied: threadpoolctl>=2.0.0
Requirement already satisfied: joblib>=0.11 in /usr/
Collecting pydeck>=0.1.dev5
  Downloading pydeck-0.7.1-py2.py3-none-any.whl (4.3
    |████████████████████████████████████████| 4.3 MB 61.4
Collecting pympler>=0.9
  Downloading Pympler-1.0.1-py3-none-any.whl (164 kB
    |████████████████████████████████████████| 164 kB 20.5
Requirement already satisfied: semver in /usr/local/
Requirement already satisfied: tornado>=5.0 in /usr/
Requirement already satisfied: cachetools>=4.0 in /us
Collecting blinker>=1.0.0
  Downloading blinker-1.5-py2.py3-none-any.whl (12 k
Requirement already satisfied: tzlocal>=1.1 in /usr/
Collecting watchdog
  Downloading watchdog-2.1.9-py3-none-manylinux2014_
    |████████████████████████████████████████| 78 kB 8.1 MB
Requirement already satisfied: pillow>=6.2.0 in /usr,
```

```

Requirement already satisfied: tomli in /usr/local/lib/python3.11.0
Collecting rich>=10.11.0
  Downloading rich-12.5.1-py3-none-any.whl (235 kB)
    |████████████████████████████████████████| 235 kB 71.9 MB/s
Requirement already satisfied: altair>=3.2.0 in /usr/local/lib/python3.11.0
Collecting validators>=0.2
  Downloading validators-0.20.0.tar.gz (30 kB)
Requirement already satisfied: jsonschema>=3.0 in /usr/local/lib/python3.11.0
Requirement already satisfied: jinja2 in /usr/local/lib/python3.11.0
Requirement already satisfied: toolz in /usr/local/lib/python3.11.0
Requirement already satisfied: entrypoints in /usr/local/lib/python3.11.0
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1 in /usr/local/lib/python3.11.0
Requirement already satisfied: importlib-resources>=5.10.0 in /usr/local/lib/python3.11.0
Requirement already satisfied: ipywidgets>=7.0.0 in /usr/local/lib/python3.11.0
Requirement already satisfied: ipykernel>=5.1.2 in /usr/local/lib/python3.11.0
Requirement already satisfied: traitlets>=4.3.2 in /usr/local/lib/python3.11.0
Requirement already satisfied: jupyter-client in /usr/local/lib/python3.11.0
Requirement already satisfied: ipython>=5.0.0 in /usr/local/lib/python3.11.0
Requirement already satisfied: pexpect in /usr/local/lib/python3.11.0
Collecting jedi>=0.10
  Downloading jedi-0.18.1-py2.py3-none-any.whl (1.6 MB)
    |████████████████████████████████████████| 1.6 MB 59.9 MB/s
Requirement already satisfied: prompt-toolkit<2.1.0,>=2.0.0 in /usr/local/lib/python3.11.0
Requirement already satisfied: decorator in /usr/local/lib/python3.11.0
Requirement already satisfied: pickleshare in /usr/local/lib/python3.11.0
Requirement already satisfied: backcall in /usr/local/lib/python3.11.0
Requirement already satisfied: pygments in /usr/local/lib/python3.11.0
Requirement already satisfied: widgetsnbextension~=4.0.0 in /usr/local/lib/python3.11.0
Requirement already satisfied: jupyterlab-widgets~=3.0.0 in /usr/local/lib/python3.11.0
Requirement already satisfied: parso<0.9.0,>=0.8.0 in /usr/local/lib/python3.11.0
Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.11.0
Requirement already satisfied: wcwidth in /usr/local/lib/python3.11.0

```

!pip install transformers

```

Looking in indexes: https://pypi.org/simple, https://pypi.org/simple
Collecting transformers
  Downloading transformers-4.21.1-py3-none-any.whl (4.7 MB)
    |████████████████████████████████████████| 4.7 MB 20.8 MB/s
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11.0
Collecting tokenizers!=0.11.3,<0.13,>=0.11.1
  Downloading tokenizers-0.12.1-cp37-cp37m-manylinux_2_17_x86_64.whl (6.6 MB)
    |████████████████████████████████████████| 6.6 MB 58.8 MB/s
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11.0
Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.11.0
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.11.0
Requirement already satisfied: filelock in /usr/local/lib/python3.11.0
Requirement already satisfied: requests in /usr/local/lib/python3.11.0
Collecting huggingface-hub<1.0,>=0.1.0
  Downloading huggingface_hub-0.8.1-py3-none-any.whl (101 kB)
    |████████████████████████████████████████| 101 kB 13.5 MB/s
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.11.0
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.11.0
Requirement already satisfied: typing-extensions>=3.7.4 in /usr/local/lib/python3.11.0
Requirement already satisfied: pyparsing!=3.0.5,>=2.4.2 in /usr/local/lib/python3.11.0
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.11.0
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.11.0
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.11.0
Requirement already satisfied: urllib3!=1.25.0,!=1.26.0,>=1.21.1 in /usr/local/lib/python3.11.0
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11.0

```

```
Installing collected packages: tokenizers, huggingface-hub, transformers
Successfully installed huggingface-hub-0.8.1 tokenizers-0.10.2 transformers-4.11.3
Installing collected packages: jedi, urllib3, smmap,
```

Gerekli kütüphane ve metrikler tanımlandı

```
Uninstalling urllib3-1.24.3:
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

import tensorflow as tf
import torch
from sklearn.preprocessing import LabelEncoder
from transformers import BertTokenizer,AdamW , BertForSequenceClassification
from torch.utils.data import TensorDataset, random_split,
from sklearn.model_selection import StratifiedKFold
from sklearn.metrics import balanced_accuracy_score, class_report

import datetime
import random
import seaborn as sns
import numpy as np
import time
import matplotlib.pyplot as plt

from sklearn.metrics import f1_score
from sklearn.metrics import recall_score
from sklearn.metrics import precision_score
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score
```

'dbmdz/bert-base-turkish-128k-uncased': vocab size -> 128k.
The final training corpus has a size of 35GB and 44,04,976,662 tokens.

Temizlenmiş verinin kategorik etiketleri bir data frame e kaydedilip numerik olarak değiştirildi. Veri içeriğinde yer alan duplicate kayıtlar temizlendi .

```
# Dataset
df = pd.read_csv('/content/clean_data.csv')
df['label'] = LabelEncoder().fit_transform(df['label'])

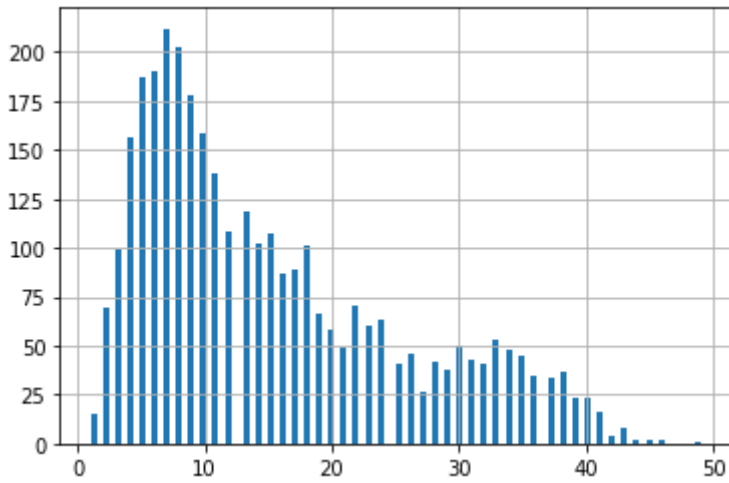
df.drop_duplicates(subset='clean_data',inplace=True)
```

Kelime sayısı sağa çarpık . Dağılımın geneline uygun bir max length seçmek gerekirse 100 alabiliriz.

```
seq_len = [len(i.split()) for i in df["clean_data"]]

pd.Series(seq_len).hist(bins = 100) #kelime sayısı sağa ç
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f53a8b2c



önceden eğitilmiş sınıflandırıcılar download edildi .

```
from simpletransformers.classification import Classification
from sklearn.model_selection import KFold
from transformers import logging
logging.set_verbosity_error()

tokenizer = BertTokenizer.from_pretrained('dbmdz/bert-base
model = BertForSequenceClassification.from_pretrained('dbn
# https://simpletransformers.ai/docs/classification-specif
```

GPU kullanıyor mu ? Memory size nedir ?

```
!nvidia-smi
```

Fri Aug 19 18:00:10 2022

```
+-----+
| NVIDIA-SMI 460.32.03      Driver Version: 460.32.03
|-----+-----+
| GPU   Name           Persistence-M| Bus-Id        Disp
| Fan   Temp    Perf    Pwr:Usage/Cap|               Memory-Usa
|-----+-----+
|  0    Tesla T4             Off      | 00000000:00:04.0 O
| N/A   43C        P0      26W / 70W   | 284MiB / 15109M
+-----+-----+
```

```
+-----+
| Processes:
| GPU    GI    CI          PID    Type    Process name
|          ID    ID
|=====|
+-----+
```

```
# check GPU
device_name = tf.test.gpu_device_name()
if device_name == '/device:GPU:0':
    device = torch.device("cuda") #Current GPU;tensörleri
else:
    raise SystemError('GPU device not found')
```

clean datadaki her satır için encode edilmiş değerleri hesaplar.
input_ids -> verilen token pozisyonunun gerçek token içerip içermediğini veya sıfır dolgulu bir konum olup olmadığını gösterir.
attention_mask yığındaki örneklerin uzunlukları farklı olsa bile, transformatöre bir yığın göndermemize olanak tanır.

0'ları dolgu jetonlarının konumlarına ve 1'leri gerçek jetonların konumlarına yerleştirerek hangi input_ids lerin dummy olduğunu belirler.

max_length=10 olsun . [101, 2026, 2171, 2003, 11754, 102, 0, 0, 0, 0], 101:[CLS]ve 102:[SEP] olduğu ifadeye karşılık gelir.

encode_plus-> encode-dan farkı tokenizer ve kelimeleri kullanarak bir string-i, bir dizi id e çevirir, modelin iki diziyi ayırt etmesine izin veren bir tür maskeleme bekler

```
input_ids = []
attention_masks = []
max_len = 100

for text in df['clean_data']:
    encoded_dict = tokenizer.encode_plus(
        str(text),
        add_special_tokens = True,
        max_length = max_len,
        pad_to_max_length = True,
        return_attention_mask = True,
        return_tensors = 'pt',
    )

    input_ids.append(encoded_dict['input_ids'])
    attention_masks.append(encoded_dict['attention_mask'])

input_ids = torch.cat(input_ids, dim=0)#Verilen boyutta(0)
attention_masks = torch.cat(attention_masks, dim=0)
```

```
print('Original: ', df['clean_data'][0])
print('Token IDs:', input_ids[0])
```

[illegible]

Stratified K fold kullanılarak farklı etiketlerin verideki dağılımları dikkate alınmış oldu, veri seti 10 a bölündü . Veri artık input_ids ve labels isimli tensörlerden oluşmakta. Tensör nesneleri ile oluşan data frame sınıflandırıcıya verildi.

```
skf = StratifiedKFold(n_splits=10)
results = []
train_data = pd.DataFrame(columns=['texts', 'labels'])
train_data['texts'] = df.clean_data
train_data['labels'] = df.label
for train_index, val_index in skf.split(train_data['texts']):
    # splitting Dataframe (dataset not included)
    train_df = train_data.iloc[train_index]
    val_df = train_data.iloc[val_index]
    # Defining Model
    model = ClassificationModel('bert', 'dbmdz/bert-base-t
                                args={'reprocess_input_data':
                                      'num_train_epochs':
                                      "output_dir": "bert_
                                      "load_best_model_at_
                                      "max_length": 30, "to

    # train the model
    model.train_model(train_df)
    # validate the model
    result, model_outputs, wrong_predictions = model.eval_
    print(f'Acc : {result["acc"]}')
    # append model score
    results.append(result["acc"])

print("results", results)
print(f"Mean-Precision: {sum(results) / len(results)}")
```



```
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
/usr/local/lib/python3.7/dist-packages/transformers/
FutureWarning,
/usr/local/lib/python3.7/dist-packages/torch/optim/lr
"https://pytorch.org/docs/stable/optim.html#how-to
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
Acc : 0.7850746268656716
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
/usr/local/lib/python3.7/dist-packages/transformers/
FutureWarning,
/usr/local/lib/python3.7/dist-packages/torch/optim/lr
"https://pytorch.org/docs/stable/optim.html#how-to
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
Acc : 0.9343283582089552
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
/usr/local/lib/python3.7/dist-packages/transformers/
FutureWarning,
/usr/local/lib/python3.7/dist-packages/torch/optim/lr
"https://pytorch.org/docs/stable/optim.html#how-to
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
Acc : 0.8776119402985074
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
/usr/local/lib/python3.7/dist-packages/transformers/
FutureWarning,
/usr/local/lib/python3.7/dist-packages/torch/optim/lr
"https://pytorch.org/docs/stable/optim.html#how-to
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
Acc : 0.9101796407185628
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
/usr/local/lib/python3.7/dist-packages/transformers/
FutureWarning,
/usr/local/lib/python3.7/dist-packages/torch/optim/lr
"https://pytorch.org/docs/stable/optim.html#how-to
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
Acc : 0.8473053892215568
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
/usr/local/lib/python3.7/dist-packages/transformers/
FutureWarning,
/usr/local/lib/python3.7/dist-packages/torch/optim/lr
"https://pytorch.org/docs/stable/optim.html#how-to
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
Acc : 0.8562874251497006
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
/usr/local/lib/python3.7/dist-packages/transformers/
FutureWarning,
/usr/local/lib/python3.7/dist-packages/torch/optim/lr
"https://pytorch.org/docs/stable/optim.html#how-to
```

```

/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to pandas
Acc : 0.8592814371257484
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to pandas
/usr/local/lib/python3.7/dist-packages/transformers/
FutureWarning,
/usr/local/lib/python3.7/dist-packages/torch/optim/
"https://pytorch.org/docs/stable/optim.html#how-to-
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to pandas
Acc : 0.8053892215568862
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to pandas
/usr/local/lib/python3.7/dist-packages/transformers/
FutureWarning,
/usr/local/lib/python3.7/dist-packages/torch/optim/
"https://pytorch.org/docs/stable/optim.html#how-to-
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to pandas
Acc : 0.8832335329341318
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to pandas
/usr/local/lib/python3.7/dist-packages/transformers/
FutureWarning,
/usr/local/lib/python3.7/dist-packages/torch/optim/
"https://pytorch.org/docs/stable/optim.html#how-to-
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to pandas
Acc : 0.8622754491017964
results [0.7850746268656716, 0.9343283582089552, 0.87
Mean-Precision: 0.8620967021181517

```

```
#model.get_named_parameters()
```

```

predictions = model_outputs.argmax(axis=1)
actuals = val_df.labels.values

```

```

print(classification_report(actuals, predictions, digits=3)
print(balanced_accuracy_score(actuals, predictions))
cm = confusion_matrix(actuals, predictions, labels=[0,1,2,3])
print(cm)
tn, fp, fn, tp = cm
print((tn, fp, fn, tp))

```

	precision	recall	f1-score	support
0	0.942	0.831	0.883	59
1	0.840	0.875	0.857	41
2	0.802	0.775	0.789	89
3	0.877	0.928	0.901	134

```

accuracy          0.862      33.4
macro avg         0.865      0.852      0.858      33.4
weighted avg      0.863      0.862      0.862      33.4

0.8520814013341834
[[ 49   1   7   2]
 [  0  42   4   2]
 [  1   5  69  14]
 [  2   2   6 128]]
(array([49,  1,  7,  2]), array([ 0, 42,  4,  2]), array([ 1,  5, 69, 14], dtype=object), array([ 2,  2,  6, 128], dtype=object))

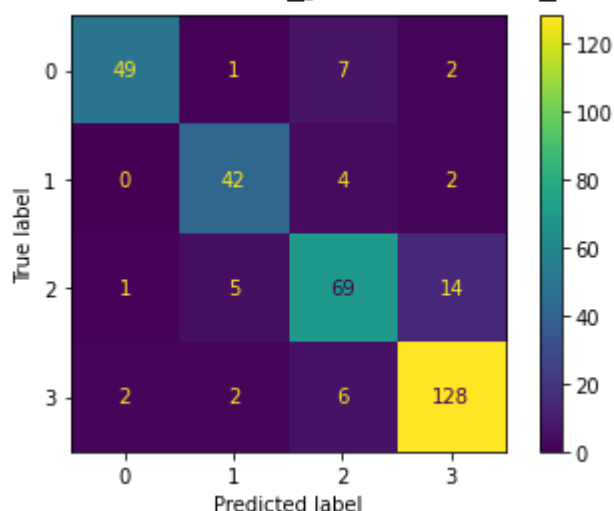
```

```

disp = ConfusionMatrixDisplay(confusion_matrix=cm,
                              display_labels=[0,1,2,3])
disp.plot()

```

```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay>
```



Double-click (or enter) to edit

```

skf = StratifiedKFold(n_splits=10)
results = []
train_data = pd.DataFrame(columns=['texts', 'labels'])
train_data['texts'] = df.clean_data
train_data['labels'] = df.label
for train_index, val_index in skf.split(train_data['texts'], train_data['labels']):
    # splitting Dataframe (dataset not included)
    train_df = train_data.iloc[train_index]
    val_df = train_data.iloc[val_index]
    # Defining Model
    model = ClassificationModel('bert', 'dbmdz/bert-base-tiny',
                                args={'reprocess_input_data': True,
                                      'num_train_epochs': 10,
                                      "output_dir": "bert_output",
                                      "load_best_model_at_end": True,
                                      "max_length": 100, "tokenizer": "bert-base-tiny"})
    # train the model
    model.train_model(train_df)
    # validate the model
    results.append(model.validate_model(val_df))

```

```
    result, model_outputs, wrong_predictions = model.eval_
    print(f'Acc : {result["acc"]}')
# append model score
    results.append(result["acc"])

print("results",results)
print(f"Mean-Precision: {sum(results) / len(results)}")
```



```
disp.plot()
```

	precision	recall	f1-score	support
0	0.943	0.847	0.893	50
1	0.860	0.896	0.878	43
2	0.809	0.809	0.809	89
3	0.901	0.928	0.914	128
accuracy			0.877	330
macro avg	0.878	0.870	0.873	330
weighted avg	0.878	0.877	0.877	330

```
0.8699539890952449
```

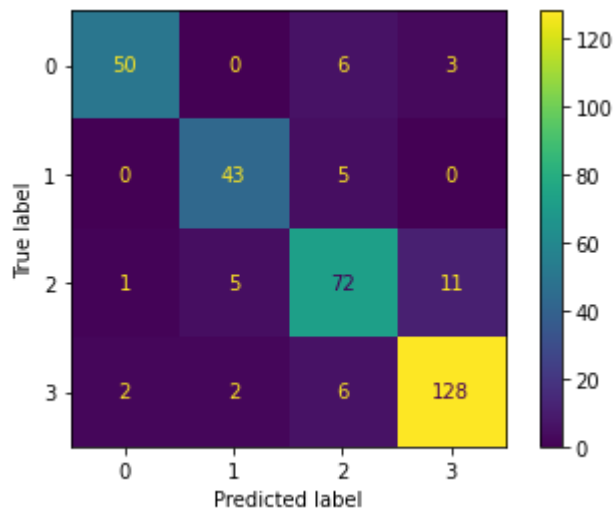
```
[[ 50  0  6  3]
```

```
 [ 0 43  5  0]
```

```
 [ 1  5 72 11]
```

```
 [ 2  2  6 128]]
```

```
(array([50,  0,  6,  3]), array([ 0, 43,  5,  0]), as
<sklearn.metrics._plot.confusion_matrix.ConfusionMat
```



```
skf = StratifiedKFold(n_splits=10)
results = []
train_data = pd.DataFrame(columns=['texts', 'labels'])
train_data['texts'] = df.clean_data
train_data['labels'] = df.label
for train_index, val_index in skf.split(train_data['texts']
    # splitting Dataframe (dataset not included)
    train_df = train_data.iloc[train_index]
    val_df = train_data.iloc[val_index]
    # Defining Model
    model = ClassificationModel('bert', 'dbmdz/bert-base-t
        args={'reprocess_input_data': True,
              'num_train_epochs': 10,
              'output_dir': "bert_
              "load_best_model_at_end": True,
              "max_length": 100, "t

    # train the model
    model.train_model(train_df)
    "
```

```
# validate the model
    result, model_outputs, wrong_predictions = model.eval_
    print(f'Acc : {result["acc"]}')
# append model score
    results.append(result["acc"])

print("results",results)
print(f"Mean-Precision: {sum(results) / len(results)}")
```

```

/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
/usr/local/lib/python3.7/dist-packages/transformers/
FutureWarning,
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_
"https://pytorch.org/docs/stable/optim.html#how-to-
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
Acc : 0.7970149253731343
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
/usr/local/lib/python3.7/dist-packages/transformers/
FutureWarning,
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_
"https://pytorch.org/docs/stable/optim.html#how-to-
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
Acc : 0.9223880597014925
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
/usr/local/lib/python3.7/dist-packages/transformers/
FutureWarning,
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_
"https://pytorch.org/docs/stable/optim.html#how-to-
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
Acc : 0.9373134328358209
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
/usr/local/lib/python3.7/dist-packages/transformers/
FutureWarning,
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_
"https://pytorch.org/docs/stable/optim.html#how-to-
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
Acc : 0.9191616766467066
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
/usr/local/lib/python3.7/dist-packages/transformers/
FutureWarning,
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_
"https://pytorch.org/docs/stable/optim.html#how-to-
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
Acc : 0.8892215568862275
/usr/local/lib/python3.7/dist-packages/simpletransformers/
>Dataframe headers not specified. Falling back to 1
/usr/local/lib/python3.7/dist-packages/transformers/

```

```

predictions = model_outputs.argmax(axis=1)
actuals = val_df.labels.values

```

```

print(classification_report(actuals, predictions, digits=3))
print(balanced_accuracy_score(actuals, predictions))
cm = confusion_matrix(actuals, predictions, labels=[0,1,2])
print(cm)
tn, fp, fn, tp = cm
print((tn, fp, fn, tp))

```



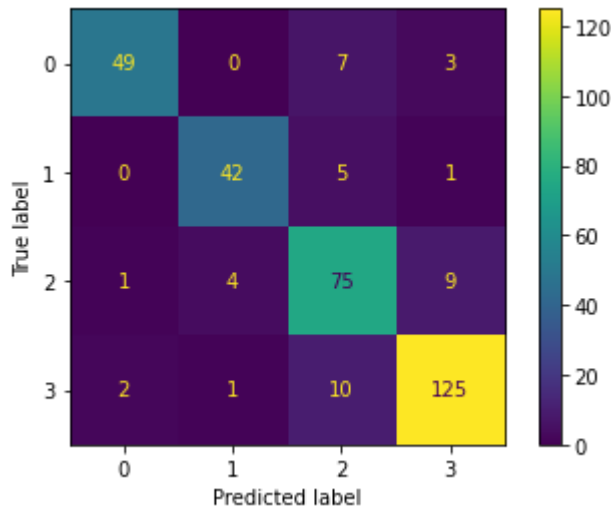
```
disp = ConfusionMatrixDisplay(confusion_matrix=cm,
                              display_labels=[0,1,2,3])
disp.plot()
```

	precision	recall	f1-score	support
0	0.942	0.831	0.883	59
1	0.894	0.875	0.884	48
2	0.773	0.843	0.806	84
3	0.906	0.906	0.906	131
accuracy			0.871	332
macro avg	0.879	0.864	0.870	332
weighted avg	0.875	0.871	0.872	332

```
0.8635005513097574
```

```
[[ 49  0  7  3]
 [ 0 42  5  1]
 [ 1  4 75  9]
 [ 2  1 10 125]]
```

```
(array([49, 0, 7, 3]), array([ 0, 42, 5, 1]), array([ 1, 4, 75, 9],
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay object>
```



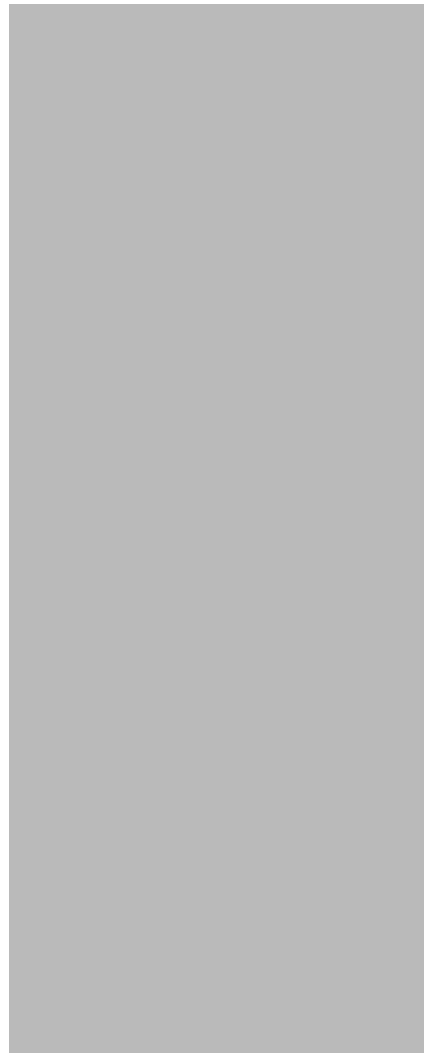
Bir sonraki adım şöyle olabilirdi ;

Multi-Label Classification Çok etiketli metin sınıflandırmasında, veri kümesinden tek bir örneğin hedefi, n farklı ikili etiketlerin bir listesidir. Bir transformatör tabanlı çok etiketli metin sınıflandırma modeli, tipik olarak, üzerinde bir sınıflandırma katmanı bulunan bir transformatör modelinden oluşur. Sınıflandırma katmanı, n sayıda her etikete karşılık gelen çıkış nöronlarına sahip olacaktır. Her çıktı nöronu (ve uzantı olarak, her etiket) birbirinden bağımsız olarak kabul edilir.

<https://simpletransformers.ai/docs/multi-label-classification/>

BERT modelinde parametre sayısı nasıl hesaplanır?

<https://stackoverflow.com/questions/64485777/how-is-the-number-of-parameters-be-calculated-in-bert-model>



✓ 0s completed at 10:53 PM

● ✕