

# pdf-summarization

October 13, 2024

```
[1]: !pip install PyPdf2
      !pip install transformers
      !pip install --upgrade transformers
      !pip install torch
```

Collecting PyPdf2

Downloading pypdf2-3.0.1-py3-none-any.whl.metadata (6.8 kB)

Downloading pypdf2-3.0.1-py3-none-any.whl (232 kB)

232.6/232.6 kB

4.5 MB/s eta 0:00:00

Installing collected packages: PyPdf2

Successfully installed PyPdf2-3.0.1

Requirement already satisfied: transformers in /usr/local/lib/python3.10/dist-packages (4.44.2)

Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from transformers) (3.16.1)

Requirement already satisfied: huggingface-hub<1.0,>=0.23.2 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.24.7)

Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (1.26.4)

Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from transformers) (24.1)

Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (6.0.2)

Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (2024.9.11)

Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from transformers) (2.32.3)

Requirement already satisfied: safetensors>=0.4.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.4.5)

Requirement already satisfied: tokenizers<0.20,>=0.19 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.19.1)

Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.10/dist-packages (from transformers) (4.66.5)

Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub<1.0,>=0.23.2->transformers) (2024.6.1)

Requirement already satisfied: typing-extensions>=3.7.4.3 in

```

/usr/local/lib/python3.10/dist-packages (from huggingface-
hub<1.0,>=0.23.2->transformers) (4.12.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests->transformers) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-
packages (from requests->transformers) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests->transformers) (2.2.3)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests->transformers)
(2024.8.30)
Requirement already satisfied: transformers in /usr/local/lib/python3.10/dist-
packages (4.44.2)
Collecting transformers
  Downloading transformers-4.45.2-py3-none-any.whl.metadata (44 kB)
                                44.4/44.4 kB
2.1 MB/s eta 0:00:00
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-
packages (from transformers) (3.16.1)
Requirement already satisfied: huggingface-hub<1.0,>=0.23.2 in
/usr/local/lib/python3.10/dist-packages (from transformers) (0.24.7)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-
packages (from transformers) (1.26.4)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.10/dist-packages (from transformers) (24.1)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-
packages (from transformers) (6.0.2)
Requirement already satisfied: regex!=2019.12.17 in
/usr/local/lib/python3.10/dist-packages (from transformers) (2024.9.11)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-
packages (from transformers) (2.32.3)
Requirement already satisfied: safetensors>=0.4.1 in
/usr/local/lib/python3.10/dist-packages (from transformers) (0.4.5)
Collecting tokenizers<0.21,>=0.20 (from transformers)
  Downloading tokenizers-0.20.1-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_
x86_64.whl.metadata (6.7 kB)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.10/dist-
packages (from transformers) (4.66.5)
Requirement already satisfied: fsspec>=2023.5.0 in
/usr/local/lib/python3.10/dist-packages (from huggingface-
hub<1.0,>=0.23.2->transformers) (2024.6.1)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/usr/local/lib/python3.10/dist-packages (from huggingface-
hub<1.0,>=0.23.2->transformers) (4.12.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests->transformers) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-
packages (from requests->transformers) (3.10)

```

```

Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests->transformers) (2.2.3)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests->transformers)
(2024.8.30)
Downloading transformers-4.45.2-py3-none-any.whl (9.9 MB)
          9.9/9.9 MB
45.0 MB/s eta 0:00:00
Downloading
tokenizers-0.20.1-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl
(3.0 MB)
          3.0/3.0 MB
56.9 MB/s eta 0:00:00
Installing collected packages: tokenizers, transformers
  Attempting uninstall: tokenizers
    Found existing installation: tokenizers 0.19.1
    Uninstalling tokenizers-0.19.1:
      Successfully uninstalled tokenizers-0.19.1
  Attempting uninstall: transformers
    Found existing installation: transformers 4.44.2
    Uninstalling transformers-4.44.2:
      Successfully uninstalled transformers-4.44.2
Successfully installed tokenizers-0.20.1 transformers-4.45.2
Requirement already satisfied: torch in /usr/local/lib/python3.10/dist-packages
(2.4.1+cu121)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-
packages (from torch) (3.16.1)
Requirement already satisfied: typing-extensions>=4.8.0 in
/usr/local/lib/python3.10/dist-packages (from torch) (4.12.2)
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages
(from torch) (1.13.3)
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-
packages (from torch) (3.3)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages
(from torch) (3.1.4)
Requirement already satisfied: fsspec in /usr/local/lib/python3.10/dist-packages
(from torch) (2024.6.1)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.10/dist-packages (from jinja2->torch) (2.1.5)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in
/usr/local/lib/python3.10/dist-packages (from sympy->torch) (1.3.0)

```

```

[2]: import PyPDF2
import warnings
from transformers import BertTokenizer, EncoderDecoderModel,
↳EncoderDecoderConfig
warnings.filterwarnings('ignore')

```

```
[3]: def extract_text_from_pdf(pdf_path):
    pdf_reader = PyPDF2.PdfReader(pdf_path)
    text = ""
    for page in pdf_reader.pages:
        text += page.extract_text()
    return text
```

```
[4]: model_ckpt = 'ardavey/bert2gpt-indosum'
tokenizer = BertTokenizer.from_pretrained(model_ckpt)
tokenizer.bos_token = tokenizer.cls_token
tokenizer.eos_token = tokenizer.sep_token

config = EncoderDecoderConfig.from_pretrained(model_ckpt)
config.early_stopping = True

model = EncoderDecoderModel.from_pretrained(model_ckpt, config=config)
```

```
tokenizer_config.json: 0%|          | 0.00/1.40k [00:00<?, ?B/s]
```

```
vocab.txt: 0%|          | 0.00/262k [00:00<?, ?B/s]
```

```
special_tokens_map.json: 0%|          | 0.00/182 [00:00<?, ?B/s]
```

```
config.json: 0%|          | 0.00/5.31k [00:00<?, ?B/s]
```

```
model.safetensors: 0%|          | 0.00/1.05G [00:00<?, ?B/s]
```

EncoderDecoderModel has generative capabilities, as ``prepare_inputs_for_generation`` is explicitly overwritten. However, it doesn't directly inherit from ``GenerationMixin``. From v4.50 onwards, ``PreTrainedModel`` will NOT inherit from ``GenerationMixin``, and this model will lose the ability to call ``generate`` and other related functions.

- If you're using ``trust_remote_code=True``, you can get rid of this warning by loading the model with an auto class. See

[https://huggingface.co/docs/transformers/en/model\\_doc/auto#auto-classes](https://huggingface.co/docs/transformers/en/model_doc/auto#auto-classes)

- If you are the owner of the model architecture code, please modify your model class such that it inherits from ``GenerationMixin`` (after ``PreTrainedModel``, otherwise you'll get an exception).

- If you are not the owner of the model architecture class, please contact the model code owner to update it.

```
generation_config.json: 0%|          | 0.00/320 [00:00<?, ?B/s]
```

```
[5]: def summarize_text(text):
    inputs = tokenizer(text, return_tensors="pt", max_length=512,
        ↪truncation=True)

    summary_ids = model.generate(
        inputs['input_ids'],
        max_length=150,
```

```

        num_beams=4,
        early_stopping=True
    )

    summary = tokenizer.decode(summary_ids[0], skip_special_tokens=True)
    return summary

```

```

[6]: text = """
Perubahan iklim merupakan salah satu tantangan terbesar yang dihadapi oleh
    ↳ dunia saat ini. Dampaknya dirasakan oleh setiap negara, tanpa memandang
    ↳ batas geografis. Fenomena seperti kenaikan suhu global, mencairnya es di
    ↳ kutub, dan meningkatnya frekuensi bencana alam seperti banjir dan
    ↳ kekeringan, menjadi bukti nyata dari ancaman ini. Selain itu, perubahan
    ↳ iklim juga mengancam kelangsungan hidup berbagai spesies, termasuk manusia,
    ↳ karena mempengaruhi ketersediaan pangan dan air bersih.

Pemerintah di seluruh dunia terus berupaya untuk mengatasi masalah ini melalui
    ↳ berbagai kebijakan dan perjanjian internasional. Salah satu langkah penting
    ↳ yang telah diambil adalah Perjanjian Paris, yang bertujuan untuk menurunkan
    ↳ emisi gas rumah kaca dan menjaga peningkatan suhu global di bawah 2 derajat
    ↳ Celsius. Namun, keberhasilan upaya ini sangat bergantung pada keseriusan dan
    ↳ kerjasama semua negara.

Di tingkat individu, setiap orang juga memiliki peran penting dalam menjaga
    ↳ lingkungan. Mulai dari mengurangi penggunaan plastik sekali pakai, beralih
    ↳ ke energi terbarukan, hingga menanam pohon, semua tindakan ini dapat
    ↳ membantu mengurangi dampak perubahan iklim. Pendidikan dan kesadaran akan
    ↳ pentingnya menjaga lingkungan juga perlu ditingkatkan, agar setiap orang
    ↳ memahami bahwa bumi ini adalah rumah kita bersama dan harus dijaga untuk
    ↳ generasi mendatang.
"""

summary = summarize_text(text)
print(summary)

```

perubahan iklim merupakan salah satu tantangan terbesar yang dihadapi oleh dunia saat ini. dampaknya dirasakan oleh setiap negara, tanpa memandang batas geografis. fenomena seperti kenaikan suhu global, mencairnya es di kutub, dan meningkatnya frekuensi bencana alam seperti banjir dan kekeringan, menjadi bukti nyata dari ancaman ini. selain itu, perubahan iklim juga mengancam kelangsungan hidup berbagai spesies, termasuk manusia, karena mempengaruhi ketersediaan pangan dan air bersih.

```

[8]: pdf_path = '/content/Testing-Text-Suma-1.pdf'

extract_pdf_text = extract_text_from_pdf(pdf_path)
summary = summarize_text(extract_pdf_text)

```

```
print(summary)
```

pengenalan lingkungan alam pada siswa sd memiliki banyak manfaat penting bagi perkembangan mereka, baik secara beregu, af sertifikasi, maupun psikomotor. siswa dapat mengembangkan kesadaran tentang keberadaan dan pentingnya menjaga lingkungan. berikut adalah beberapa alasan mengapa mengenalkan lingkungan alam sekitar itu penting : meningkatkan kecerdasan dan pengetahuan pengalaman langsung dengan alam membantu siswa belajar tentang berbagai macam tumbuhan, hewan, dan fenomena alam lainnya.